

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2004, 10:25:14 ; Search time 14.4172 Seconds  
(without alignments)

2831.539 Million cell updates/sec

Title: US-10-028-952a-10

Sequence: 1 IRHEGERTXEVXNKEAPG.....DXAIALQPKERETLSQKX 784

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	138.5	3.7	591	1	ALU1_HUMAN
2	126	3.4	591	1	ALU8_HUMAN
3	125	3.3	587	1	ALU2_HUMAN
4	123.5	3.3	593	1	ALU6_HUMAN
5	122	3.3	587	1	ALU3_HUMAN
6	110	2.9	2167	1	SHK1_RAT
7	108	2.9	593	1	ALU7_HUMAN
8	104	2.8	660	1	YHJ1_EBV
9	100.5	2.7	585	1	ALU5_HUMAN
10	95	2.5	705	1	SYN1_HUMAN
11	93	2.5	450	1	A2AA_PIG
12	93	2.5	609	1	NMB1_HUMAN
13	92.5	2.5	1235	1	RMT1_RABIT
14	92	2.5	971	1	AREA_GIBBU
15	89.5	2.4	1219	1	WIS1_YEAST
16	88	2.4	452	1	A2AA_BOVIN
17	88	2.3	786	1	SYN1_BOVIN
18	87	2.3	308	1	ROG_MOUSE
19	87	2.3	670	1	SYN1_MOUSE
20	86	2.3	391	1	ROG_HUMAN
21	86	2.3	450	1	A2AA_CANPO
22	86	2.3	450	1	A2AA_MOUSE
23	86	2.3	2424	1	CCAA_RABIT
24	85	2.3	833	1	SRG2_MOUSE
25	84	2.2	424	1	S3B4_HUMAN
26	84	2.2	514	1	VE2_HPV5B
27	84	2.2	870	1	SRG2_HUMAN
28	83	2.2	415	1	SYN1_CANPA
29	82.5	2.2	2483	1	PCX_DROME
30	82	2.2	704	1	SYN1_RAT
31	81.5	2.2	448	1	ACG2_DICDI
32	81.5	2.2	696	1	JPH2_MOUSE
33	81.5	2.2	1076	1	VE38_SCHPO

#### ALIGNMENTS

RESULT 1  
ALU1\_HUMAN STANDARD; PRT; 591 AA.  
ID ALU1\_HUMAN  
AC P39188;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 01-FEB-1995 (Rel. 31, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Alu subfamily J sequence contamination warning entry.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP MEDLINE=95021756; PubMed=7935834;  
RA Claverie J.-M.; Makalowski W.;  
RT "Alu alert.";  
RL Nature 371:752-752(1994).  
RN [2]  
RP CONCEPT.  
RX MEDLINE=92241891; PubMed=1572661;  
RA Claverie J.-M.;  
RT "Identifying coding exons by similarity search: alu-derived and other  
potentially misleading protein sequences.";  
RL Genomics 12:838-841(1992).  
RN [3]  
RP ALU FAMILIES CLASSIFICATION.  
RX MEDLINE=88333009; PubMed=3138422;  
RA Quentin Y.;  
RT "The Alu family developed through successive waves of fixation  
closely connected with primate lineage history.";  
RL J. Mol. Evol. 27:194-202(1988).  
RN [4]  
RP ALU FAMILIES CLASSIFICATION.  
RX MEDLINE=91178815; PubMed=1706781;  
RA Jurka J.; Milosavljevic A.;  
RT "Reconstruction and analysis of human Alu genes.";  
RL J. Mol. Evol. 32:105-121(1991).  
RN [5]  
RP MISCELLANEOUS: VARIOUS ANALYSES (SEE REF. 3 AND REF. 4) INDICATE  
THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
SENSITIVITY SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
REPEATS.  
RN [6]  
RP MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
CODON. 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.  
RN [7]  
RP CAUTION: This Alu entry is provided in order to avoid the further  
pollution of protein sequence databases with Alu-derived amino  
acid sequences.  
RN [8]  
RP CAUTION: Alu repetitive sequences are interspersed in human and  
primate genomes with an average spacing of 4 kb. Some of them are  
actively transcribed by pol. III. Normal transcripts may contain  
Alu-derived sequences in 5' or 3' untranslated regions. However,  
cDNA libraries also contain partial and/or rearranged cDNAs  
ligated with Alu-derived sequence in any orientation. Although Alu  
elements (especially situated on the complementary strand) have a  
great potential to create additional/alternative exons.

```
CC consideration should be given to the possibility that the presence
CC of an Alu in an open reading frame may have resulted from a
CC cloning artifact or may be due to misinterpretation of sequencing
CC data. This point has been overlooked on several occasions, with
CC the consequence of erroneous Alu-derived amino acid sequences
CC being reported.
CC -1- CAUTION: Any significant similarity of a putative protein sequence
CC with an Alu-translated entry must be taken as a warning that a
CC part of Alu repeat may have been artifactually included in the
CC coding nucleotide sequence.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U14567; -; NOT_ANNOTATED_CDS.
CC -----
CC Hypothetical protein.
CC FT DOMAIN 1 96 FRAME-1.
CC FT DOMAIN 100 195 FRAME-2.
CC FT DOMAIN 199 294 FRAME-3.
CC FT DOMAIN 298 393 FRAME-4.
CC FT DOMAIN 397 492 FRAME-5.
CC FT DOMAIN 496 591 FRAME-6.
CC SO SEQUENCE 591 AA; 63790 MW; 665D395735519D95 CRC64;
CC -----
CC Query Match 3.7%; Score 138.5; DB 1; Length 591;
CC Best Local Similarity 27.6%; Pred. No. 4.1e-05;
CC Matches 48; Conservative 4; Mismatches 23; Indels 99; Gaps 2;
CC -----
CC 707 LESPDPQPOHGETLSLQKNTXIX----- 731
CC Db 121 LSPDPQPOHGETLSLQKNTXIXLQKLAGRGARLXSQLGLRQEDRLSPGRCSEPSRH 180
CC -----
CC 732 ----- 731
CC Db 181 CTPWATERDVPVSKKXXPGAVAHACNPSTLGGGRITXAGFEETSLGNIVKPLRYKY 240
CC -----
CC Oy 732 ---PYLHTGSLSYSEGGKXXIDXAQVEAAAVXDXDALPLQPGKERETLSQK 782
CC Db 241 KMKGVVARRACSPSYSGKGRRLAKAEVDAVSR-DRATLALPGSGSETLSQK 293
CC -----
CC RESULT 2
CC ALU8_HUMAN STANDARD; PRT; 591 AA.
CC ID ALU8_HUMAN STANDARD; PRT; 591 AA.
CC AC P3195;
CC DT 01-FEB-1995 (Rel. 31, Created)
CC DT 01-FEB-1995 (Rel. 31, Last sequence update)
CC DT 10-OCT-2003 (Rel. 42, Last annotation update)
CC DB Alu subfamily SX sequence contamination warning entry.
CC OS Homo sapiens (Human).
CC CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
CC CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CC CC NCBI_TaxId=9606;
CC CC [1]
CC CC SEQUENCE FROM N.A.
CC CC MEDLINE=95021758; PubMed=7935834;
CC CC Claverie J.-M.; Makalowski W.;
CC CC "Alu alert.";
CC CC Nature 371:752-752(1994).
CC CC [2]
CC CC CONCEPT.
CC CC MEDLINE=92241891; PubMed=1572661;
CC CC Claverie J.-M.;
CC CC "Identifying coding exons by similarity search: alu-derived and other
CC CC potentially misleading protein sequences.";
CC CC Genomics 12:838-841(1992).
CC CC [3]
CC CC ALU FAMILIES CLASSIFICATION.
CC -----
CC RX MEDLINE=88333009; PubMed=3138422;
CC RA Quantin Y.;
CC RT "The Alu family developed through successive waves of fixation
CC RT closely connected with primate lineage history.";
CC RT J. Mol. Evol. 27:194-202(1988).
CC RL [4]
CC RN ALU FAMILIES CLASSIFICATION.
CC RP MEDLINE=91178815; PubMed=1706781;
CC RX Jurka J., Milosavljevic A.;
CC RA "Reconstruction and analysis of human Alu genes.";
CC RL J. Mol. Evol. 32:105-121(1991).
CC CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE
CC CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
CC CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
CC CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
CC CC REPEATS.
CC CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
CC CC CODON. 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
CC CC -1- CAUTION: This Alu entry is provided in order to avoid the further
CC CC pollution of protein sequence databases with Alu-derived amino
CC CC acid sequences.
CC CC -1- CAUTION: Alu repetitive sequences are interspersed in human and
CC CC primate genomes with an average spacing of 4 kb. Some of them are
CC CC actively transcribed by pol III. Normal transcripts may contain
CC CC Alu-derived sequences in 5' or 3' untranslated regions. However,
CC CC cDNA libraries also contain partial and/or rearranged cDNAs
CC CC ligated with Alu-derived sequence in any orientation. Although Alu
CC CC elements (especially situated on the complementary strand) have a
CC CC great potential to create additional/alternative exons,
CC CC consideration should be given to the possibility that the presence
CC CC of an Alu in an open reading frame may have resulted from a
CC CC cloning artifact or may be due to misinterpretation of sequencing
CC CC data. This point has been overlooked on several occasions, with
CC CC the consequence of erroneous Alu-derived amino acid sequences
CC CC being reported.
CC CC -1- CAUTION: Any significant similarity of a putative protein sequence
CC CC with an Alu-translated entry must be taken as a warning that a
CC CC part of Alu repeat may have been artifactually included in the
CC CC coding nucleotide sequence.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL; U14574; -; NOT_ANNOTATED_CDS.
CC CC -----
CC CC Hypothetical protein.
CC CC FT DOMAIN 1 96 FRAME-1.
CC CC FT DOMAIN 100 195 FRAME-2.
CC CC FT DOMAIN 199 294 FRAME-3.
CC CC FT DOMAIN 298 393 FRAME-4.
CC CC FT DOMAIN 397 492 FRAME-5.
CC CC FT DOMAIN 496 591 FRAME-6.
CC CC SO SEQUENCE 591 AA; 64395 MW; AC8154AD8A6B280 CRC64;
CC -----
CC Query Match 3.4%; Score 126; DB 1; Length 591;
CC Best Local Similarity 26.3%; Pred. No. 0.00058;
CC Matches 50; Conservative 5; Mismatches 33; Indels 102; Gaps 3;
CC -----
CC Oy 691 LIPVILALXKXAGSLSEVRDPOXQHGFTLSLQKNTXIX----- 731
CC Db 108 LTPVIPALMEED---HLRSGVDRDQPGHGETPSLTKIQKLAGRGARLXSQLGLRLOEN 164
CC -----
CC Oy 732 ----- 731
CC Db 165 RLNPGGGGSEPSRHCTPAATERDSYSKKXXRAEAGRGSGSLRSGHFGKRTXGQEF 224
CC -----
CC Oy 732 ---PYLHTGSLSYSEGGKXXIDXAQVEAAAVXDXDALPLQPGKERETLSQK 772
CC Db 225 ETSLANVPRLLYKXKXKPGVAVARACNPSPSYSGKGRRIATREAEVAVSR-DRATLALP 283
```



QY 773 GXERETLSQK 782  
 Db 284 GROSERPSQK 293

RESULT 3  
 ID ALU2\_HUMAN STANDARD; PRT; 587 AA.  
 AC P39189;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Alu subfamily 5B sequence contamination warning entry.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95021758; PubMed=7935834;  
 RA Claverie J.-M., Makalowski W.;  
 RT "Alu alert."  
 RL Nature 371:752-752(1994).  
 RN [2]  
 RP CONCEPT.  
 RX MEDLINE=92241891; PubMed=1572661;  
 RA Claverie J.-M.;  
 RT "Identifying coding exons by similarity search: alu-derived and other  
 RT potentially misleading protein sequences."  
 RL Genomics 12:838-841(1992).  
 RN [3]  
 RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=88333009; PubMed=3138422;  
 RA Quentin Y.;  
 RT "The Alu family developed through successive waves of fixation  
 RT closely connected with primate lineage history."  
 RL J. Mol. Evol. 27:194-202(1988).  
 RN [4]  
 RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=91178815; PubMed=1706781;  
 RA Jurka J., Miliusavljovic A.;  
 RT "Reconstruction and analysis of human Alu genes."  
 RL J. Mol. Evol. 32:105-121(1991).  
 CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE  
 CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
 CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
 CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
 CC REPEATS.  
 CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
 CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.  
 CC -1- CAUTION: This Alu entry is provided in order to avoid the further  
 CC pollution of protein sequence databases with Alu-derived amino  
 CC acid sequences.  
 CC -1- CAUTION: Alu repetitive sequences are interspersed in human and  
 CC primate genomes with an average spacing of 4 Kb. Some of them are  
 CC actively transcribed by pol III. Normal transcripts may contain  
 CC Alu-derived sequences in 5' or 3' untranslated regions. However,  
 CC cDNA libraries also contain partial and/or rearranged cDNAs  
 CC ligated with Alu-derived sequence in any orientation. Although Alu  
 CC elements (especially situated on the complementary strand) have a  
 CC great potential to create additional/alternative exons,  
 CC consideration should be given to the possibility that the presence  
 CC of an Alu in an open reading frame may have resulted from a  
 CC cloning artifact or may be due to misinterpretation of sequencing  
 CC data. This point has been overlooked on several occasions, with  
 CC the consequence of erroneous Alu-derived amino acid sequences  
 CC being reported.  
 CC -1- CAUTION: Any significant similarity of a putative protein sequence  
 CC with an Alu-translated entry must be taken as a warning that a  
 CC part of Alu repeat may have been artifactually included in the  
 CC coding nucleotide sequence.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

CC -----  
 CC EMBL; U14568; NOT ANNOTATED\_CDS.  
 CC KW Hypothetical protein.  
 CC FT DOMAIN 1 96 FRAME-1.  
 CC FT DOMAIN 100 194 FRAME-2.  
 CC FT DOMAIN 198 292 FRAME-3.  
 CC FT DOMAIN 296 391 FRAME-4.  
 CC FT DOMAIN 395 489 FRAME-5.  
 CC FT DOMAIN 493 587 FRAME-6.  
 CC SQ SEQUENCE 587 AA; 63703 MW; 3EAB3E3E3929203 CRC64;

Query Match 3.3%; Score 125; DB 1; Length 587;  
 Best Local Similarity 43.4%; Pred. No. 0.00071;  
 Matches 43; Conservative 7; Mismatches 33; Indels 16; Gaps 3;

QY 691 LIPVILALXXKXAGRSLSRVRDQPKHGHTL-----SLQNTKIXPYVLAHTGSLSY 743  
 Db 6 LTPVLPALMEAEAGSGSGGEI-----ETILANTVKPRLYKXKXPGVAVAGSPSY 57

QY 744 SEGKXHXIDKAGVEAAVVRXDXAIALQPKXERETLSQK 782  
 Db 58 SGGXGRWPAATREAEIIVSR-DRATVALQPGQSETPSQK 95

RESULT 4  
 ID ALU6\_HUMAN STANDARD; PRT; 593 AA.  
 AC P39193;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Alu subfamily 5P sequence contamination warning entry.  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95021758; PubMed=7935834;  
 RA Claverie J.-M., Makalowski W.;  
 RT "Alu alert."  
 RL Nature 371:752-752(1994).  
 RN [2]  
 RP CONCEPT.  
 RX MEDLINE=92241891; PubMed=1572661;  
 RA Claverie J.-M.;  
 RT "Identifying coding exons by similarity search: alu-derived and other  
 RT potentially misleading protein sequences."  
 RL Genomics 12:838-841(1992).  
 RN [3]  
 RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=88333009; PubMed=3138422;  
 RA Quentin Y.;  
 RT "The Alu family developed through successive waves of fixation  
 RT closely connected with primate lineage history."  
 RL J. Mol. Evol. 27:194-202(1988).  
 RN [4]  
 RP ALU FAMILIES CLASSIFICATION.  
 RX MEDLINE=91178815; PubMed=1706781;  
 RA Jurka J., Miliusavljovic A.;  
 RT "Reconstruction and analysis of human Alu genes."  
 RL J. Mol. Evol. 32:105-121(1991).  
 CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF.3 AND REF.4) INDICATE  
 CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
 CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
 CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU



Best Local Similarity 44.6%; Pred. No. 0.0013;  
Matches 41; Conservative 8; Mismatches 41; Indels 2; Gaps 2;

OY 691 LIPVIALXXAGSLERVDQXQHGSETSLQKNTKIXVLAHTCSLSYSGGXK 750  
DB 6 LTPVLPALWEAEAGSGRGOEITIPAKTVKP-RLYKRYKNXPGVAGACSPSLGXGR 64  
OY 751 IDAAQVEAAVAXDXAIALPOXRESETL50K 782  
DB 65 MAXTRAEALAVSR-DEPITALQORQSETP50K 95

RESULT 6  
SHKL RAT  
ID SHKL RAT STANDARD; PRT; 2167 AA.  
AC 09WV48; 09WV28; 09WV13; 09WV08;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 28-FEB-2003 (Rel. 41, Last annotation update)  
DE SH3 and multiple ankyrin repeat domains protein 1 (Shank1) (GKAP/SAPAP  
DE interacting protein) (SPANK-1) (Symanon) (Somatostatin receptor  
DE interacting protein) (SSTR interacting protein) (SSTRIP).  
GN SHANK1.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
ON NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3), AND INTERACTION WITH DLGAP1 AND  
RP DLG4.  
RC TISSUE=Brain;  
RX MEDLINE=994119021; PubMed=10488079;  
RA Yao I., Hata Y., Hizo K., Deguchi M., Ide N., Takeuchi M., Takai Y.,  
RT "Symanon, a novel neuronal protein interacting with synapse-associated  
RL protein 90/postsynaptic density-95-associated protein.";  
RL J. Biol. Chem. 274:27463-27466(1999).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 4), AND INTERACTION WITH DLGAP1.  
RC STRAIN=Sprague-Dawley;  
RX MEDLINE=99360650; PubMed=10433268;  
RA Naibit S., Kim E., Tu J.C., Xiao B., Sala C., Valschanoff J.,  
RT Weisberg R.J., Morley P.F., Sheng M.;  
RT "Shank, a novel family of postsynaptic density proteins that binds to  
RL the NMDA receptor/PSD-95/GKAP complex and cortactin.";  
RL Neuron 23:569-582(1999).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Brain;  
RX MEDLINE=20549637; PubMed=10958799;  
RA Tobben S., Suedhof T.C., Stahl B.;  
RT "The G protein-coupled receptor Ctl1 interacts directly with proteins  
RT of the Shank family.";  
RL J. Biol. Chem. 275:36204-36210(2000).  
RN [4]  
RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3; 4 AND 5), AND  
RP DEVELOPMENTAL STAGE.  
RC TISSUE=Brain;  
RX MEDLINE=99436166; PubMed=10506216;  
RA Lim S., Naibit S., Yoon U., Hwang J.I., Suh P.G., Sheng M., Kim E.;  
RT "Characterization of the Shank family of synaptic proteins. Multiple  
RT genes, alternative splicing, and differential expression in brain and  
RT development.";  
RL J. Biol. Chem. 274:29510-29518(1999).  
RN [5]  
RP PARTIAL SEQUENCE FROM N.A. (ISOFORM 4).  
RC TISSUE=Brain;  
RX MEDLINE=20020275; PubMed=10551867;  
RA Zitzer H., Heenck H.-H., Baechner D., Richter D., Kretenkamp H.-U.;  
RT "Somatostatin receptor interacting protein defines a novel family of  
RT multidomain proteins present in human and rodent brain.";  
RL J. Biol. Chem. 274:33297-33301(1999).  
RN [6]  
RP INTERACTION WITH HOMER-1, AND SUBCELLULAR LOCATION.

RX MEDLINE=99360651; PubMed=10433269;  
RA Tu J.C., Xiao B., Naibit S., Yuan J.P., Petralia R.S., Brakeman P.,  
RA Doan A., Akalu V.K., Lianhan A.A., Sheng M., Morley P.F.;  
RT "Coupling of mGluR/Homer and PSD-95 complexes by the Shank family of  
RT postsynaptic density proteins.";  
RL Neuron 23:583-592(1999).  
RN [7]  
RP INTERACTION WITH SPTAN1.  
RX MEDLINE=21523912; PubMed=11509555;  
RA Bockers T.M., Mameza M.G., Kreutz M.R., Bockmann J., Weise C.,  
RA Buck F., Richter D., Gundelfinger E.D., Kretenkamp H.-U.;  
RT "Synaptic scaffolding proteins in rat brain. Ankyrin repeats of the  
RT multidomain Shank protein family interact with the cytoskeletal  
RT protein alpha-fodrin.";  
RL J. Biol. Chem. 276:40104-40112(2001).  
RN [8]  
RP FUNCTION.  
RX MEDLINE=21389514; PubMed=11498055;  
RA Sala C., Plech V., Wilson N.R., Pasafaro M., Liu G., Sheng M.;  
RT "Regulation of dendritic spine morphology and synaptic function by  
RT Shank and Homer.";  
RL Neuron 31:115-130(2001).  
RN [9]  
RP REVIEW.  
RX MEDLINE=20267867; PubMed=10860936;  
RA Sheng M., Kim E.;  
RL J. Cell Sci. 113:1851-1856(2000).  
CC -1- FUNCTION. Seems to be an adaptor protein in the postsynaptic  
CC density (PSD) of excitatory synapses that interconnects receptors  
CC of the postsynaptic membrane including NMDA-type and metabotropic  
CC glutamate receptors, and the actin-based cytoskeleton. May play a  
CC role in the structural and functional organization of the  
CC dendritic spine and synaptic junction. Overexpression promotes  
CC maturation of dendritic spines and the enlargement of spine heads  
CC via its ability to recruit Homer to postsynaptic sites, and  
CC enhances presynaptic function.  
CC -1- SUBUNIT: May homodimerize via its SAM domain. Interacts with  
CC SPTAN1, Homer-1 and DLGAP1/GKAP. Is part of a complex with  
CC DLG4/PSD-95 and DLGAP1/GKAP. Interacts with SSTR2 C-terminus via  
CC the PDZ domain (By similarity).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic; postsynaptic density of  
CC neuronal cells. Colocalizes with alpha-latrotoxin receptor 1.  
CC -1- ALTERNATIVE PRODUCTS:  
CC Event=Alternative splicing; Named isoforms=5;  
CC Comment=Additional isoforms seem to exist;  
CC Name=1;  
CC IsoId=Q9WV48-1; Sequence=Displayed;  
CC Name=2;  
CC IsoId=Q9WV48-2; Sequence=VSP\_006072; VSP\_006073;  
CC Name=3;  
CC IsoId=Q9WV48-3; Sequence=VSP\_006074;  
CC Name=4; Synonyms=A;  
CC IsoId=Q9WV48-4; Sequence=VSP\_006075;  
CC Name=5;  
CC IsoId=Q9WV48-5; Sequence=VSP\_006076; VSP\_006077;  
CC -1- TISSUE SPECIFICITY: Expressed only in brain (neurofil of cortex,  
CC CAL region hippocampus and molecular layer of cerebellum).  
CC -1- DEVELOPMENTAL STAGE: Expression increases from low levels at birth  
CC to high levels at 3-4 weeks before dropping slightly in adulthood.  
CC Expressed in the cortex and the molecular layer of the cerebellum  
CC at postnatal day 7. Isoform 2 expression does not change during  
CC development of both cortex and cerebellum. Isoform 4 expression  
CC decreases significantly during development of cortex but not  
CC cerebellum.  
CC -1- SIMILARITY: BELONGS TO THE SHANK FAMILY.  
CC -1- SIMILARITY: Contains 7 ANK repeats.  
CC -1- SIMILARITY: Contains 1 PDZ/DHR domain.  
CC -1- SIMILARITY: Contains 1 sterile alpha motif (SAM) domain.  
CC -1- SIMILARITY: Contains 1 SH3 domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -



```

CC -----
CC primate genomes with an average spacing of 4 kb. Some of them are
CC actively transcribed by pol III. Normal transcripts may contain
CC Alu-derived sequences in 5' or 3' untranslated regions. However,
CC cDNA libraries also contain partial and/or rearranged cDNAs
CC related with Alu-derived sequence in any orientation. Although Alu
CC elements (especially situated on the complementary strand) have a
CC great potential to create additional/alternative exons,
CC consideration should be given to the possibility that the presence
CC of an Alu in an open reading frame may have resulted from a
CC cloning artifact or may be due to misinterpretation of sequencing
CC data. This point has been overlooked on several occasions, with
CC the consequence of erroneous Alu-derived amino acid sequences
CC being reported.
CC -1- CAUTION: Any significant similarity of a putative protein sequence
CC with an Alu-translated entry must be taken as a warning that a
CC part of Alu repeat may have been artifactually included in the
CC coding nucleotide sequence.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U14573; -, NOT_ANNOTATED_CDS.
CC KW Hypothetical protein.
CC FT DOMAIN 1 97 FRAME-1.
CC FT DOMAIN 101 196 FRAME-2.
CC FT DOMAIN 200 295 FRAME-3.
CC FT DOMAIN 239 395 FRAME-4.
CC FT DOMAIN 399 484 FRAME-5.
CC FT DOMAIN 488 593 FRAME-6.
CC SQ SEQUENCE 593 AA; 64417 MW; 54A4F50F33A6089F CRC64;
CC -----
Query Match 2.8%; Score 108; DB 1; Length 593;
Best Local Similarity 24.6%; Pred. No. 0.027;
Matches 43; Conservative 7; Mismatches 25; Indels 100; Gaps 3;
CC -----
QY 707 LESRVDPQXHGELSLQKNTKIX----- 731
DB 122 IFSGVDPQGHGTPPLTKIQKLAGGGGRLXSQLRLRQENRLNPGGGGCEPERSRH 181
QY 732 ----- 731
DB 182 CTPANATRAKLLKXXXXGAVAHANCPETLGGRGWITXGGEFTSLANNVKKRLYYKY 241
QY 732 ---PYVLHTGSLSTSEKGGXIDYAOEVEAAAVKDXAIALQPG-XEREFTLSQK 782
DB 242 KXPGVAVAGACNPYSYSGXGRIRAXTREAEVAVSR-DRATLALPQGOQERNVSXK 295
CC -----

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; V01555; -, NOT_ANNOTATED_CDS.
CC DR PIR; A03742; OQHE3.
CC KW Hypothetical protein; Early protein; Repeat.
CC FT DOMAIN 149 648 4 X 125 AA TANDM REPEATS.
CC FT REPEAT 149 273 1.
CC FT REPEAT 274 398 2.
CC FT REPEAT 399 523 3.
CC FT REPEAT 524 648 4.
CC SQ SEQUENCE 660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;
CC -----
Query Match 2.8%; Score 104; DB 1; Length 660;
Best Local Similarity 23.2%; Pred. No. 0.07;
Matches 87; Conservative 13; Mismatches 103; Indels 172; Gaps 21;
CC -----
QY 18 APGAP--PVSPRGARGGRRPCG-----PVKYHYSDRXXTDVRGGEPR- 61
DB 188 APGAPGPAP--QPGGAAVPSGATPHERGSGADPPA-----AARLPFERQ--EPRL 238
QY 62 -GALAKAKRPARRRPAATPSGASARWGYLGGVYQOTXQVSGELREDN-----LPWX 116
DB 239 PQDLAAQRCPCAG--PEPTSGAA-----QTHRRPPGCPRSANRNGCCRTYR 285
QY 117 RRAKARLILFTWTDXESGASRFPFGEXAXXVKVTGTLGLRPSXHSDAFXSPD 176
DB 286 RR-----SGAQGHPPPG-----AQGRSGPTG--GRPAAPG----- 315
QY 177 VGSSYKXAEFTKRWIVHPLIGKXSNVYTVRQVSTLXMCCHGNPQAYENRXXRH 236
DB 316 -----APGTPA----- 321
QY 237 VVYLKAGNAGKXXSVGLXLNASKSESPPGTIRQRGASVGLGXPPRLSP----- 288
DB 322 ---APPGGGAIVPG---ATPHPERGSGPADPPAARLPERQEPRLPQDLAAQRC 373
QY 289 PAGRPPPTRXKRAAGSVPRRAPG-----PSSXACPSNSW-----TERGK 330
DB 374 PAG--PPPTSGAAAGQTHRRPPGCPRSANPG---CPRTWRRSQAQGRHPPGAGOR 427
QY 331 ---GGXPLARRAP 340
DB 428 PSCPTGGRPAAPAP 442
CC -----

```

```

RESULT 8
YH11_EBV STANDARD; PRT; 660 AA.
ID YH11_EBV
AC P03181;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Hypothetical BHLF1 protein.
OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84370667; PubMed=6087149;
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tufnell P.S., Barrett B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211 (1984).

```

```

RESULT 9
ALUS_HUMAN STANDARD; PRT; 585 AA.
ID ALUS_HUMAN
AC P39152;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Alu subfamily SC sequence contamination warning entry.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95021758; PubMed=7935834;
RA Claverie J.-M., Makalowski W.;
RT "Alu alert.";
RL Nature 371:752-752 (1994).
RN [2]
RP CONCEPT.
RX MEDLINE=92241891; PubMed=1572661;

```

RA Clavette J.-M.;  
 RT "Identifying coding exons by similarity search: alu-derived and other  
 RT potentially misleading protein sequences.";  
 RL Genomics 12:838-841(1992).  
 RN [3]  
 RN ALU FAMILIES CLASSIFICATION.  
 RP MEDLINE=88333009; PubMed=3138422;  
 RA Quenec'h Y.;  
 RT "The Alu family developed through successive waves of fixation  
 RT closely connected with primate lineage history.";  
 RL J. Mol. Evol. 27:194-202(1988).  
 RN [4]  
 RN ALU FAMILIES CLASSIFICATION.  
 RP MEDLINE=91178815; PubMed=1706781;  
 RA Jurka J., Mironavich A.;  
 RT "Reconstruction and analysis of human Alu genes.";  
 RL J. Mol. Evol. 32:105-121(1991).  
 CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF. 3 AND REF. 4) INDICATE  
 CC THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING  
 CC CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX  
 CC FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU  
 CC REPEATS.  
 CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP  
 CC CODON. 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.  
 CC -1- CAUTION: This Alu entry is provided in order to avoid the further  
 CC pollution of protein sequence databases with Alu-derived amino  
 CC acid sequences.  
 CC -1- CAUTION: Alu repetitive sequences are interspersed in human and  
 CC primate genomes with an average spacing of 4 kb. Some of them are  
 CC actively transcribed by pol III. Normal transcripts may contain  
 CC Alu-derived sequences in 5' or 3' untranslated regions. However,  
 CC cDNA libraries also contain partial and/or rearranged cDNAs  
 CC ligated with Alu-derived sequence in any orientation. Although Alu  
 CC elements (especially situated on the complementary strand) have a  
 CC great potential to create additional/alternative exons,  
 CC consideration should be given to the possibility that the presence  
 CC of an Alu in an open reading frame may have resulted from a  
 CC cloning artifact or may be due to misinterpretation of sequencing  
 CC data. This point has been overlooked on several occasions, with  
 CC the consequence of erroneous Alu-derived amino acid sequences  
 CC being reported.  
 CC -1- CAUTION: Any significant similarity of a putative protein sequence  
 CC with an Alu-translated entry must be taken as a warning that a  
 CC part of Alu repeat may have been attractually included in the  
 CC coding nucleotide sequence.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: U14571; -; NOT ANNOTATED\_CDS.  
 KW Hypothetical protein.  
 FT DOMAIN 1 95 FRAME-1.  
 FT DOMAIN 1 99 193 FRAME-2.  
 FT DOMAIN 197 291 FRAME-3.  
 FT DOMAIN 295 389 FRAME-4.  
 FT DOMAIN 393 487 FRAME-5.  
 FT DOMAIN 491 585 FRAME-6.  
 SQ SEQUENCE 585 AA; 63957 MM; 46E88C4F493650A7 CRC64;  
 Query Match 2.7%; Score 100.5; DB 1; Length 585;  
 Best Local Similarity 41.4%; Pred. NO. 0.13; Index 17; Gaps 4;  
 Matches 41; Conservative 6; Mismatches 35;  
 CY 691 LIPVILALXXAKGRSLERVRDQPKXHGELT-----SLQKNTKIXPYVLAHTCSLSY 743  
 DB 6 LTPVILPLWEMKAGSGSGEI-----ETILANVVKPRLYKXKXGLGVAVRACSPSY 57  
 CY 744 SGGXGXIXDAQGEVLAANVXKDXAIALQPKXERTLSQK 782

Db 58 SGGXGXIRIXTREAFAVAVSR-DEATLQPG-DRAFLRLX 94  
 RESULT 10  
 SYN1 HUMAN  
 ID SYN1 HUMAN STANDARD; PRT; 705 AA.  
 AC P17600; O75825;  
 DT 01-AUG-1990 (Rel. 15, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Synapsin I (Brain protein 4.1).  
 GN SYN1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=90243651; PubMed=2110562;  
 RA Suedhof T.C.;  
 RT "The structure of the human synapsin I gene and protein.";  
 RL J. Biol. Chem. 265:7849-7852(1990).  
 RN [2]  
 RP SEQUENCE OF 1-125 FROM N.A.  
 RX MEDLINE=90368667; PubMed=2118519;  
 RA Sauerwald A., Hoese C., Oeschwald R., Kiliann M.W.;  
 RT "The 5'-flanking region of the synapsin I gene. A G-C-rich, TATA- and  
 RT CAAT-less, phylogenetically conserved sequence with cell  
 RT type-specific promoter function.";  
 RL J. Biol. Chem. 265:14932-14937(1990).  
 RN [3]  
 RP SEQUENCE OF 1-258 FROM N.A.  
 RA Gieffem D.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Neuronal phosphoprotein that coats synaptic vesicles,  
 CC binds to the cytoskeleton, and is believed to function in the  
 CC regulation of neurotransmitter release.  
 CC -1- SUBCELLULAR LOCATION: Synapse.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Synapsin 1A;  
 CC Name=Synapsin 1A;  
 CC IsoId=P17600.1; Sequence=Displayed;  
 CC Name=Synapsin 1B;  
 CC IsoId=P17600.2; Sequence=VSP 006316, VSP 006317;  
 CC -1- PTM: SUBSTRATE OF AT LEAST FOUR DIFFERENT PROTEIN KINASES. IT IS  
 CC PROBABLE THAT PHOSPHORYLATION PLAYS A ROLE IN THE REGULATION OF  
 CC SYNAPSIN I IN THE NERVE TERMINAL.  
 CC -1- SIMILARITY: Belongs to the synapsin family.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M58378; AAC41930.1; -;  
 CC EMBL: M58321; AAC41930.1; JOINED.  
 CC EMBL: M58341; AAC41930.1; JOINED.  
 CC EMBL: M58351; AAC41930.1; JOINED.  
 CC EMBL: M58353; AAC41930.1; JOINED.  
 CC EMBL: M58359; AAC41930.1; JOINED.  
 CC EMBL: M58371; AAC41930.1; JOINED.  
 CC EMBL: M58372; AAC41930.1; JOINED.  
 CC EMBL: M58373; AAC41930.1; JOINED.  
 CC EMBL: M58374; AAC41930.1; JOINED.  
 CC EMBL: M58375; AAC41930.1; JOINED.  
 CC EMBL: M58376; AAC41930.1; JOINED.  
 CC EMBL: M58377; AAC41930.1; JOINED.  
 CC EMBL: M58378; AAC41931.1; ALT\_SEQ.

```

DR EMBL, M58321, AAC41931.1, JOINED.
DR EMBL, M58341, AAC41931.1, JOINED.
DR EMBL, M58351, AAC41931.1, JOINED.
DR EMBL, M58359, AAC41931.1, JOINED.
DR EMBL, M58371, AAC41931.1, JOINED.
DR EMBL, M58372, AAC41931.1, JOINED.
DR EMBL, M58373, AAC41931.1, JOINED.
DR EMBL, M58374, AAC41931.1, JOINED.
DR EMBL, M58375, AAC41931.1, JOINED.
DR EMBL, M58376, AAC41931.1, JOINED.
DR EMBL, M58377, AAC41931.1, JOINED.
DR EMBL, M58301, AAC60508.1, -.
DR EMBL, AL009172, CAAL557.1, -.
DR PIR, A35363, A35363.
DR HSSP, P17599, IADU.
DR GENEW, HGNC:11494, SYN1.
DR MIM, 313440, -.
DR GO, GO:0005803, C:secretory vesicle; TAS.
DR GO, GO:0005215, F:transporter activity; TAS.
DR GO, GO:0006899, P:nonselective vesicle transport; TAS.
DR GO, GO:0007268, P:synaptic transmission; TAS.
DR InterPro, IPR01359, Synapsin.
DR Pfam, PF02078, Synapsin, 1.
DR Pfam, PF02750, Synapsin, C, 1.
DR PRINTS, PR01368, SYNAPSIN.
DR PROSITE, PS00415, SYNAPSIN_1, 1.
DR PROSITE, PS00416, SYNAPSIN_2, 1.
DR Synapse, Phosphorylation; Neurone, Repeat; Actin-binding;
Alternative splicing.
FT DOMAIN 1 28 A.
FT DOMAIN 29 112 B (LINKER).
FT DOMAIN 113 420 C
(Actin-binding AND SYNAPTIC-VESICLE
BINDING).
FT DOMAIN 421 655 D (PRO-RICH LINKER).
FT MOD_RES 9 E.
FT MOD_RES 568 568 PHOSPHORYLATION (BY PKA AND CAMK1).
FT MOD_RES 605 605 PHOSPHORYLATION (BY CAMK2).
FT VAAPLIC 661 669 NKSQSLRYNA->KASPAQAP (in isoform
Synapsin IB).
FT VAAPLIC 670 705 /FTid=VSP_006316.
FT COMPACT 138 138 Missing (in isoform Synapsin IB).
FT SEQUENCE 705 AA, 73954 MW, 487831123FF6882F CRC64;
Query Match 2.5%, Score 95, DB 1, Length 705;
Best Local Similarity 27.5%, Pred. No. 0.51;
Matches 53, Conservative 16, Mismatches 64, Indels 60, Gaps 11;
QY 177 VGSSTVHSAEFTKRWIVPLIGXKS-----WVTVVQVAFLLMCCCHGNPAQYERN 230
DB 388 VGSSN-----PLIGDHODEKQILVELVYNNKMAAL-----PRGRD 425
QY 231 RRRRLVVLGXGANGAKXSVGLXNKSSESPPXGTTIRRRASVGLGXPA-----R 265
DB 426 ASP-----GRSGHGQTPSPGALPLGRQTSQ-QPAGPPAQRPPPG-QPPQPGPPQR 476
QY 286 LSPF-AGRP-----STRXXRAGGRVRRAPAGGXAACPPSSMET-----GRGRK- 331
DB 477 QGSPFQGRPPPGGCHLSGLPPRAGSPPLQLPSTLAPQPPAQAAPPPGGRGRSPV 536
QY 332 ---GXPLARHP 340
DB 537 AGSGAPAPARP 549
RESULT 11
AAAA_PIG STANDARD; FRT; 450 AA.
ID A2AA_PIG
AC P18671;

```

```

DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Alpha-2A adrenergic receptor (Alpha-2AAR) (Alpha-2AR).
GN ADR2A OR A2AR.
OS Sue scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
OX NCBI_TaxID=9823;
[1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 212-232.
RC TISSUE=Liver;
RX MEDLINE=91009157; PubMed=2170371;
RA Guyer C.A., Horstman D.A., Wilson A.L., Clark J.D., Kravoe E.J. Jr.,
RA Limbird L.E.;
RT "Cloning, sequencing, and expression of the gene encoding the porcine
RT alpha 2-adrenergic receptor. Allosteric modulation by Na+, H+, and
RT amiloride analogs."
RL J. Biol. Chem. 265:17307-17317(1990).
RN [2]
RP MUTAGENESIS OF CYS-442.
RX MEDLINE=93216775; PubMed=8385131;
RA "Mutations of the alpha 2A-adrenergic receptor that eliminate
RT detectable palmitoylation do not perturb receptor-G-protein
RT coupling."
RL J. Biol. Chem. 268:8003-8011(1993).
CC -1- FUNCTION: Alpha-2 adrenergic receptors mediate the catecholamine-
CC induced inhibition of adenylate cyclase through the action of G
CC proteins.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- MISCELLANEOUS: Alpha2-adrenergic receptor shows an allosteric
CC modulation by Na(+), H(+) and amiloride analogs.
CC -1- SIMILARITY: Belongs to family 1 of G-protein coupled receptors.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, J05652, AAA30984.1, -.
DR PIR, A38316; A38316.
DR HSSP, P29274; 1MMH.
DR InterPro, IPR000276; GPCR_Rhodopsn.
DR Pfam, PF00001, 7tm.1, 1.
DR PRINTS, PR00237, GPCR_RHODOPSIN.
DR PROSITE, PS00237, G-PROTEIN_RECEP_F1_1, 1.
DR PROSITE, PS00262, G-PROTEIN_RECEP_F1_2, 1.
DR G-protein coupled receptor; Transmembrane; Glycoprotein;
KW Multigene family; Phosphorylation; Lipoprotein; Palmitate.
FT DOMAIN 1 33
FT TRANSMEM 34 59
FT DOMAIN 60 70
FT TRANSMEM 71 96
FT DOMAIN 97 106
FT TRANSMEM 107 129
FT DOMAIN 130 149
FT TRANSMEM 150 173
FT DOMAIN 174 192
FT TRANSMEM 193 217
FT DOMAIN 218 374
FT TRANSMEM 375 399
FT TRANSMEM 400 409
FT TRANSMEM 410 430
FT DOMAIN 431 450
FT CARBOHYD 10 14
FT CARBOHYD 14 14
FT DISULFID 106 188
FT LIPID 442 442
FT SITE 113 113

```





RT Sequence features of the regions of 1,367,185 bp covered by 19  
 RT physically assigned PI and TAC clones."  
 CC DNA Res. 5:203-216(1998).  
 CC -1- FUNCTION: Eliminates the production of nonsense-containing RNAs  
 CC (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Potential).  
 CC -1- SIMILARITY: Belongs to the DNA/NAM7 helicase family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AB013394; BAB10240.1; -  
 KM Hypohectical protein; Hydrolase; Helicase; ATP-binding; Zinc-finger.  
 FT ZN\_FING 148 178 C2H2-TYPE (ATYPICAL) (POTENTIAL).  
 FT ZN\_FING 202 232 C4-TYPE (Potential).  
 FT NP\_BIND 507 514 ATP (POTENTIAL).  
 FT DOMAIN 102 105 POLY-SER.  
 SQ SEQUENCE 1235 AA; 134870 MW; 1773F6C524D2E6A CRC64;  
 Query Match 2.5%; Score 92.5; DB 1; Length 1235;  
 Best Local Similarity 29.9%; Pred. No. 1.6;  
 Matches 44; Conservative 9; Mismatches 59; Indels 35; Gaps 9;  
 QY 226 QYERNR--XHLVYVLGXGANGAKXSVGLXNLSKSESR-----PXGTRQRRG 274  
 Db 929 QGQPKRKIYNDRLRFYGAGMIGNDNPFSG--NPADRGRSGRAGSYLPBG---PENG 984  
 QY 275 ASVGL---GXPR--LSPPAGRPPTSTXXRAGRVPRARAGPSXACEPSWETGRG 329  
 Db 985 AAPPGLPAGYPIPRVPLSPFGPPSPQPYA-----IPTRGP-VGAVPHAPQPGNHGCGA 1037  
 QY 330 KGGXFLARAPVRAAEFXSSTIHN 356  
 Db 1038 GGGTSVGHLPHQQA-----TOHN 1056  
 RESULT 14  
 AREA GIBFU STANDARD; PRT; 971 AA.  
 ID AREA GIBFU STANDARD; PRT; 971 AA.  
 AC P78688;  
 DT 15-JUL-1998 (Rel. 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 28-FEB-2003 (Rel. 41, Last annotation update)  
 DE Nitrogen regulatory protein area.  
 GN AREA.  
 OS *Gibberella fujikuroi* (Bakanease and foot rot disease fungus) (*Fusarium moniliforme*).  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Hypocreomycetidae; Hypocreales; Nectriaceae; *Gibberella*.  
 OK NCBI\_TaxID=5127;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ms67;  
 RX MEDLINE=99168774; PubMed=100721216;  
 RA Tudyanski B., Homann V., Feng B., Marzluf G.A.;  
 RT "Isolation, characterization and disruption of the area nitrogen  
 RT regulatory gene of *Gibberella fujikuroi*.";  
 RL Mol. Genet. 261:106-114(1999).  
 CC -1- FUNCTION: MAJOR NITROGEN REGULATORY PROTEIN. POSITIVELY ACTING  
 CC REGULATORY GENE OF NITROGEN METABOLITE REPRESSION (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- SIMILARITY: Contains 1 GAT-type zinc finger.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: Y11006; CA71897.1; -  
 DR HSSP; P17429; 4GAT.  
 DR InterPro; IPR000679; Znf GATA.  
 DR Pfam; PF030320; GATA; 1.  
 DR PRINTS; PR00619; GATAZNFINGER.  
 DR SMART; SM00401; Znf GATA; 1.  
 DR PROSITE; PS00344; GATA\_ZN\_FINGER\_1; 1.  
 DR PROSITE; PS01114; GATA\_ZN\_FINGER\_2; 1.  
 KM Transcription regulation; Activator; DNA-binding; Zinc-finger;  
 KM Nuclear protein; Nucleate assimilation.  
 FT ZN\_FING 694 718 GATA-TYPE.  
 SQ SEQUENCE 971 AA; 103580 MW; 887D882141C7453 CRC64;  
 Query Match 2.5%; Score 92; DB 1; Length 971;  
 Best Local Similarity 20.8%; Pred. No. 1.4;  
 Matches 76; Conservative 30; Mismatches 138; Indels 122; Gaps 18;  
 QY 6 GEETEXYXNXX--EAPGAPVSPRGAGGXRRCGSPVKYHSDRXXTDPVRGGEPRG- 62  
 Db 580 GTTDFPDNNGDMESNGLERSQGSFRGMLRQHPXL-----PRNASTVHFVGQGNRF 634  
 QY 63 -ALAXKAPRAARRPGATRSGXARMGVXKGRCTCTXXQVXSGLREDRNLPMXRARA 121  
 Db 635 EQLAQSSQSPADPDGNGTMSGFSS--VAPSR-----PSPFMSQGS- 674  
 QY 122 RLILIFSTNTDSESGASRSFPGFAXXVYRKTTGTGLMRPSKSDVAFXSFDVGSY 181  
 Db 675 -----TTNIGAAAGNND-----GNAPTTCTNCTFTPTLMRNPPSGQLCNACGLFLK 724  
 QY 182 FXEAEFTKRWIVHPLIGKXSWVTVVRQVSTFLXKCCCHGAPQYERNRKXHLVYVLG 241  
 Db 725 HG-----VVRPL-----SLKTDVIX-----KRNK----- 744  
 QY 242 XGANGAKXSVGLXNLSKSESRPXGTRIRGASVGLX-----PXPL-SP 288  
 Db 745 SGTN---VPGV--GSSTRSKTAATLNSRKNTLSMSTATNSTKPNSSNTPTVTP 797  
 QY 289 PAGRPPPS-----TRXXRAGGRVPR--APGGSXACEPSSWETGRGRGKXP 334  
 Db 798 PATSQPPSSKDVSQVSGTTSAGTASGTNSHFSGPSSGAV-----GKGWVP 848  
 QY 335 LARHAP 340  
 Db 849 IAAAP 854  
 RESULT 15  
 MYSS YEAST STANDARD; PRT; 1219 AA.  
 ID MYSS YEAST STANDARD; PRT; 1219 AA.  
 AC O04439;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Myosin-5 isoform.  
 GN MYO5 OR YMR109W OR YMR718.08.  
 OS *Saccharomyces cerevisiae* (Baker's Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; *Saccharomyces*.  
 OK NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=S288C / AB972;  
 RX MEDLINE=97313268; PubMed=9169872;  
 RA Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,  
 RA Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,  
 RA Unger K., Dye G., Moule S., Odeli C., Pearson D., Rajandream M.A.,  
 RA Rice P., Skelton J., Walsh S., Whitehead S., Batteil B.G.;  
 RT "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome  
 RT XIII.";  
 RL Nature 387:90-93(1997).



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2004, 10:45:34 ; Search time 41.7603 Seconds  
(without alignments)  
5923.479 Million cell updates/sec

Title: US-10-028-952a-10  
Perfect score: 3752  
Sequence: 1 IRHEGERTXEVXNCKEAPG.....DXAIALOPGKERETLSQKX 784

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues  
Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP\_ARCHAEA:\*  
2: SP\_BACTERIA:\*  
3: SP\_FUNGI:\*  
4: SP\_HUMAN:\*  
5: SP\_INVERTEBRATE:\*  
6: SP\_MAMMAL:\*  
7: SP\_MHC:\*  
8: SP\_ORNITHINE:\*  
9: SP\_PHAGE:\*  
10: SP\_PLANT:\*  
11: SP\_RODENT:\*  
12: SP\_VIRUS:\*  
13: SP\_VERTEBRATE:\*  
14: SP\_UNCLASSIFIED:\*  
15: SP\_VIRUS:\*  
16: SP\_BACTERIAP:\*  
17: SP\_ARCHAEP:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	119	3.2	543	16	Q82AUV7	Q82AUV7 streptomycete
2	117	3.1	95	2	Q848M1	Q848M1 bacillus me
3	113	3.0	77	16	Q8OCV6	Q8OCV6 staphylococ
4	104.5	2.8	539	10	Q9XHV3	Q9XHV3 oryza sativi
5	103	2.7	376	16	Q8XYP2	Q8XYP2 deinococcus
6	102	2.7	62	17	Q972FO	Q972FO sulfoblobus
7	99.5	2.7	655	16	Q9FBR7	Q9FBR7 streptomyces
8	99.5	2.7	674	2	Q9FIRJ6	Q9FIRJ6 acidovorax
9	97.5	2.6	377	2	Q914Z6	Q914Z6 mycobacteri
10	97.5	2.6	750	10	Q9SDPF5	Q9SDPF5 oryza sativi
11	97.5	2.6	2496	11	Q8VHND8	Q8VHND8 mus musculu
12	96.5	2.6	542	12	Q84357	Q84357 mastomys na
13	96	2.6	150	16	Q9RSH9	Q9RSH9 deinococcus
14	95.5	2.5	160	6	Q8HKB4	Q8HKB4 maccaca fasc
15	95	2.5	936	10	Q9XKX5	Q9XKX5 oryza sativ
16	94.5	2.5	409	10	Q8S1S6	Q8S1S6 oryza sativ

17	94	2.5	461	10	Q9FTD4	Q9FTD4 oryza sativ
18	93	2.5	362	10	Q8S031	Q8S031 oryza sativ
19	93	2.5	494	17	Q9HNR5	Q9HNR5 halobacteri
20	93	2.5	1305	11	Q80YF9	Q80YF9 mus musculu
21	92.5	2.5	439	10	Q8R226	Q8R226 oryza sativ
22	92.5	2.5	1243	10	Q8S3K7	Q8S3K7 arabidopsis
23	92	2.5	264	10	Q7XGB8	Q7XGB8 oryza sativ
24	91.5	2.4	437	5	Q8SXL9	Q8SXL9 drosophila
25	91.5	2.4	1627	10	Q84ZL0	Q84ZL0 oryza sativ
26	91	2.4	216	2	Q8VPR2	Q8VPR2 micrococcus
27	91	2.4	390	6	Q97GE9	Q97GE9 smitthopsis
28	90.5	2.4	272	10	Q7XGE9	Q7XGE9 oryza sativ
29	90.5	2.4	298	2	Q8EYK5	Q8EYK5 thermus the
30	90.5	2.4	437	5	Q9VWV3	Q9VWV3 drosophila
31	90.5	2.4	788	16	Q9SLS2	Q9SLS2 rhizobium
32	90	2.4	529	2	P94909	P94909 microbacter
33	90	2.4	601	5	Q62185	Q62185 caenorhabdi
34	90	2.4	3262	11	Q9EQJ5	Q9EQJ5 mus musculu
35	89.5	2.4	751	10	Q9FRM7	Q9FRM7 arabidopsis
36	89.5	2.4	759	2	Q68843	Q68843 streptomyce
37	89.5	2.4	776	10	Q8H0T8	Q8H0T8 arabidopsis
38	89.5	2.4	780	10	Q93ZT6	Q93ZT6 arabidopsis
39	89.5	2.4	780	10	Q8H179	Q8H179 arabidopsis
40	89	2.4	690	10	Q7XDU3	Q7XDU3 oryza sativ
41	88.5	2.4	219	12	Q91T11	Q91T11 tupiaia herp
42	88.5	2.4	224	17	Q9YD60	Q9YD60 aeropyrum P
43	88.5	2.4	457	10	Q94LX1	Q94LX1 oryza sativ
44	86.5	2.4	846	16	Q82FV2	Q82FV2 streptomyce
45	88	2.3	172	4	Q8NFO8	Q8NFO8 homo sapien

## ALIGNMENTS

RESULT 1

Q82AUV7 PRELIMINARY; PRT; 543 AA.

AC Q82AUV7;  
ID 01-JUN-2003 (TREMBLrel. 24, Created)  
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)  
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
DE Hypothetical protein.  
GN SAV5958.  
OS Streptomyces avermitilis.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomycinae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxID=33903;

KN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RX MEDLINE=21477403; PubMed=11572948; Hanamoto A., Takahashi C.,  
RA Omura S., Ikeda H., Ishikawa J., Hanamoto A., Shinose M.,  
RA Shinose M., Takahashi Y., Horikawa H., Nakazawa H., Osomoe T.,  
RA Kikuchi H., Shiba T., Sakaki Y., Hattori M.;  
RT "Genome sequence of an industrial microorganism Streptomyces  
RT avermitilis: deducing the ability of producing secondary  
RT metabolites.";  
RT Proc. Natl. Acad. Sci. U.S.A. 98:12215-12220(2001).

RL (2)  
RN SEQUENCE FROM N.A.  
RP STRAIN=MA-4680 / ATCC 31267 / NCIMB 12804 / NRRL 8165;  
RC MEDLINE=22608306; PubMed=12692562;  
RX Ikeda H., Ishikawa J., Hanamoto A., Shinose M., Kikuchi H., Shiba T.,  
RA Sakaki Y., Hattori M., Omura S.;  
RA "Complete genome sequence and comparative analysis of the industrial  
RT microorganism Streptomyces avermitilis.";  
RT Nat. Biotechnol. 21:526-531(2003).  
RL EMBL; AP005044; BAC73670.1; -;  
KW Hypothetical protein; Complete proteome.  
SQ SEQUENCE 543 AA; 60464 MW; E59CE37E88AD0B05 CRC64;

Query Match 3.2%; Score 119; DB 16; Length 543;  
Best Local Similarity 22.1%; Pred. No. 0.00092;

Matches 96; Conservative 41; Mismatches 145; Indels 152; Gaps 24;

```

QY 26 PRGARGXRRRCGPVTKHNS-----DXXYD-----PARGGEPGG 62
DB 121 PLREKGGRRAGG---YHSRQEWHAARLHLEDRLEQARTDRDAGVHVVRGK--- 173
QY 63 ALAAGAKRPAARRPGATSGXSARM-----GVXLRGTCQTXQVSKGE----- 106
DB 174 HLARTRHHLAAQ--LLETGWRQREAEERWFLKADGSGKRYGNETVRVSPBGEVASIKLP 231
QY 107 --LREDNLP-----KKRAKALILFTNTXESGASGSPFPGF 146
DB 232 APLKELNAPRGRVYLAQCVHFRHGOEMADREANALVRYRHDTARG--RWYLTASN 289
QY 147 XAXXVRYK---TTGITGL-----WRPSXK---SDVAFKSPDY-GSSYXEA 186
DB 290 QIPRTQVPIEALQGVIGVDMADHLAMRLDVHGNPVGDRPRFPDLSGPAHRDAQ 349
QY 187 FTKRWIVHPLIGXSW-----VXTVROVSFTLLMKCCCHNPACYEENRRLHLYVL 240
DB 350 ---VHALTGLHMAACGVTAAVEDLDFADKTRERHG-----KRRFRROLISGM 398
QY 241 GXGANGAKXXSV-----GLXINA-----SKSESRP-XGTIR--QSRGASVGL--- 279
DB 399 PTGRLRLRLTSMADHTGIAIANDPATVSRKGAQHMOKPLTSKIKRTTRDAAVAIGR 458
QY 280 --GXPRPLSPAGRPPTSTXKAGRVPR-----RAFPGSKAECPSWE 324
DB 459 AOGPIRRRT---APPHDQSDRAGHRTVQARPGIPWRBGTBRLPGPRTSRVGP--- 510
QY 325 TGRGRKGGXPLARH 338
DB 511 -GRGANAGDQNAQH 523

```

## RESULT 2

```

0848W1 PRELIMINARY; PRT; 95 AA.
AC 0848W1.
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Bacillus megaterium.
OC Plasmid pBM400.
CX Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
NX NCBI_TaxID=1404;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=QM B1551;
RX MEDLINE=21150449; PubMed=11251820;
RA Kunnimalaiyaan M., Stevenson D.M., Zhou Y., Vary P.S.;
RT "Analysis of the replication region and identification of an RNA operon
on pBM400 of Bacillus megaterium QM B1551.";
RL Mol. Microbiol. 39:1010-1021(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=QM B1551;
RA Vary P.S., Scholte M.D., White C.A., Kunnimalaiyaan M.;
RT "Complete Sequencing and Characterization of pBM400 from Bacillus
megaterium.";
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF142677; AA052805.1; -.
DR GO; GO:0046821; C:extrachromosomal DNA; IEA.
DR InterPro; IPR009019; K1.pirk.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 95 AA; 10201 MW; 3FD616C2E7379256 CRC64;

```

Query Match 3.1%; Score 117; DB 2; Length 95;  
 Best Local Similarity 41.9%; Pred. No. 0.00016;  
 Matches 31; Conservative 1; Mismatches 42; Indels 0; Gaps 0;  
 136 GASRSFXPFQXAXXVRYKTTGTLGMRPSXHSDDVAFSPVGSSTXHEAEFTKRWIVHP 195

```

DB 14 GLSDPVYHGRALIAQRKATPGITGLSPFRVHIDGFWHLIDVGSSHPGAVVPGKMAVRP 73
QY 196 LIGXSWXTVYRQ 209
DB 74 LKRYASWQVNVVRQ 87

```

## RESULT 3

```

08CQV6 PRELIMINARY; PRT; 77 AA.
AC 08CQV6.
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Conserved hypothetical protein.
SE2264.
OS Staphylococcus epidermidis.
CX Bacteria; Firmicutes; Bacillales; Staphylococcus.
NX NCBI_TaxID=1282;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 12228;
RA Zhang Y., Ren S., Li H., Fu G., Lu L., Lu G., Jia J., Tu Y., Qin Z.,
RA Chen Z., Wen Y.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB016751; AA005906.1; -.
DR InterPro; IPR009019; K1.pirk.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 77 AA; 8402 MW; 9ED757600933581A CRC64;

```

Query Match 3.0%; Score 113; DB 16; Length 77;  
 Best Local Similarity 43.3%; Pred. No. 0.00038;  
 Matches 29; Conservative 1; Mismatches 37; Indels 0; Gaps 0;

```

QY 143 PRGYAXXVRYKTTGTLGMRPSXHSDDVAFSPVGSSTXHEAEFTKRWIVHPLIGXSW 202
DB 3 PHGRAIAQRKATPGITGLSPFRVHIDGFWHLIDVGSSHPGAVVPGKMAVRP LKRYASW 62
QY 203 VXTVYRQ 209
DB 63 VQNVYRQ 69

```

## RESULT 4

```

09XHV3 PRELIMINARY; PRT; 539 AA.
AC 09XHV3.
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE 10A191.13.
GN 10A191.13.
OS Oryza sativa (Rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Shariatideae; Oryzeae; Oryza.
NX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Lemont;
RA Vysotskaya V.S., Schwartz J.R., Osborne B.I., Wing R., Yu G., Kwan A.,
RA Liu S., Lee J., Toriumi M., Luros J., Li J., Kremenetskaya I., Oji O.,
RA Theologis A.;
RT "Oryza sativa chromosome 1 BAC 10A191.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC007858; AAD39598.1; -.
DR Gramene; O9XHV3; -.
SQ SEQUENCE 539 AA; 58285 MW; D30F0686A5946605 CRC64;

```

Query Match 2.8%; Score 104.5; DB 10; Length 539;  
 Best Local Similarity 22.0%; Pred. No. 0.003;  
 Matches 95; Conservative 39; Mismatches 181; Indels 117; Gaps 17;

QY 4 EHGEEITXEVXNKCEAPGAPVPGARAGXRRPCG-PPVKYHYSDRXXTDP-----VRR 56  
 DB 120 EEDENTE-----EPPRAPIVSPRAGPRGARLADBPACH--ARPTTPRIVLVYLR 172  
 QY 57 GGEPRGALLAKGAKRPAARRPGATRGXSRMGVXLGRYTCOTXXOVXG-----105  
 DB 173 PPRPAPAAAY-TRALKKERNHARGSGKADMLAEKRCVPCNCRSGPFPFGHLSVE 231  
 QY 106 -----ELREDPMLPWKRRAKARLLIFSTNTDXESGASRFXPPFGXXXVAKV 154  
 DB 222 IFNHLPYLDVCLANTDRHIFSSRTAFELRII-----SNTYPAHWLGRRLKHT 280  
 QY 155 TTGITGLWRPSXHSVDVAFXSFDVGSSYH-----XEAFTKRWIVH-----PLIGXXSW 202  
 DB 281 DMLRRSARRMQEMEA-----TASESHLPWWLCPSPAHDCRCRTMHAHSGPVGSSYT 335  
 QY 203 VXTV-VROVSTFLLMCCCH-----GNPAQYERNRXLVYVLGANGAKXKXVGLXLN 257  
 DB 336 VLTALPPLATVSSGLGTHTRARGERGGRSRAG-----GGGDSGGGEEVGATV 389  
 QY 258 A-----SKSESXPXGTRRGRGASVGLGXFP 284  
 DB 390 AVPAGQATANLRSLASLHGLGSEERRRPREETRGDSRP--PRQRHSGARLHRRSP 447  
 QY 285 RISPAGPPPESTRXXRAGRVPRAPPGGSXAECPSSWET-GRGRKGXFLAHARVYR 343  
 DB 448 RVGAAPLGPFAARQI-----WPPRSFGPAAAVTPPGPATVAAARRQIWPFRPSPVG 502  
 QY 344 ARAEFXSGSTIH 355  
 DB 503 AAAPLGPVAVH 514

## RESULT 5

Q9RYP2 PRELIMINARY; PRT; 376 AA.  
 ID Q9RYP2  
 AC Q9RYP2  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Adenine deaminase-related protein.  
 GN DRA0268.  
 OS Deinococcus radiodurans.  
 OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;  
 OC Deinococcaceae; Deinococcus.  
 NC NCBITaxID=1259;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCIB 9279;  
 RX MEDLINE=20036896; PubMed=10567266;  
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
 RA Vamathevan J., Lam P., McDonald L., Uetrecht T., Zaleski C.,  
 RA Nakrova K.S., Aaravind L., Daly M.J., Minton K.W., Fleischman R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser C.M.;  
 RT "Genome sequence of the radioresistant bacterium Deinococcus  
 RT radiodurans R1."  
 RL Science 286:1571-1577 (1999).  
 DR EMBL; AB001863; AAF12376.1; .  
 DR TIGR; C75580; C75580.  
 DR Complete proteome.  
 SQ SEQUENCE 376 AA; 39845 MW; 7AB7FF32F8C45651 CRC64;

Query Match 2.7%; Score 103; DB 16; Length 376;

Best Local Similarity 23.0%; Pred. No. 0.031; Indels 136; Gaps 18;

Matches 82; Conservative 15; Mismatches 123; Indels 136; Gaps 18;

20 GAPVSPFGARCGXRRP--CGPPVYHYSDRXXTDPVRGGEPRGAL---AXGAKRPAAR 74

DB 64 GRPAPGAGQSGAGCEPRVYAHPALY---SGGRADPRAGSNAAGARYELLGPAPRGD 120  
 QY 75 RQATRS-GXSARMGVXIGRYTCOTXXOVXGELR-----EDRNLPP--KXRRAARL 123  
 DB 121 RFWVSRLRAARLALTLRG-----AGULRGSGSAARRRAAPPAKRRRRRAAL 169  
 QY 124 ILIFSTNTDXESGASRFXPPFGXAXKVRKVTGIGLWRPSXHSVDVAFXSFDVGSSYHX 183  
 DB 170 -----GRGLHRS-----ARALHAPDVSRPDRH-----192  
 QY 184 EAEFTKRWIVHPIIGXSWXIVVROVSTFLLMCCCHG---NEAQYERNRXLVYVL 240  
 DB 193 -----RACGAGRRRPPARRRPLRARRVVELL 219  
 QY 241 GXGANGAKXKXVGLXLNASKSESPXGTRRGRGAS-----VGIKXPPL 286  
 DB 220 DVGQPARRRP-CHQSARASGGPAR--RQRGPARGSGARTARRRHPGRGRGPPA 276  
 QY 287 SPAPGPPPESTRXXRAGRVPRAPPGGSXAECPSSWETGRGRKGXPLA--RHAP 340  
 DB 277 VAPALR--GSDDRPASGR-GRRCPPGD-----GGALAGLHIALSRHP 318

## RESULT 6

Q972F0 PRELIMINARY; PRT; 62 AA.

ID Q972F0  
 AC Q972F0  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)  
 DE Hypothetical protein STS132.  
 GN STS132.  
 OS Sulfolobus tokodaii.  
 OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;  
 OC Sulfolobus.  
 NC NCBITaxID=111955;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JCM 10545 / ?;  
 RX MEDLINE=21456156; PubMed=1572479;  
 RA Kwarabiyasi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M.,  
 RA Sekine M., Baba S.-I., Anka A., Kosugi H., Hosoyama A., Fukui S.,  
 RA Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y.,  
 RA Yoshizawa T., Tanaka T., Kudoh Y., Nakazaki J., Kuehida N., Oguchi A.,  
 RA Aoki K.-I., Masuda S., Yanagii M., Nishimura M., Yamagishi A.,  
 RA Ohima T., Kikuchi H.;  
 RT "Complete genome sequence of an aerobic thermoacidophilic  
 RT Crenarchaeon, Sulfolobus tokodaii strain7."  
 RL DNA Res. 8:123-140 (2001).  
 DR EMBL; AP000985; BAB66219.1; .  
 KM Hypothetical protein; Complete proteome.  
 SQ SEQUENCE 62 AA; 6269 MW; 134E8A389A075DD8 CRC64;

Query Match 2.7%; Score 102; DB 17; Length 62;

Best Local Similarity 48.1%; Pred. No. 0.0044; Indels 0; Gaps 0;

Matches 25; Conservative 2; Mismatches 25; Indels 0; Gaps 0;

11 RKTTPGITSSRARAHIDPAVVCYIDVSSHPGGAAPKGRARAPLMSGVSWV 62

## RESULT 7

Q9FBR7 PRELIMINARY; PRT; 655 AA.

ID Q9FBR7  
 AC Q9FBR7  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)  
 DE Putative iron chelatase.  
 GN SCO5278 OR SCB12.02.  
 OS Streptomyces coelicolor.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Streptomycinae; Streptomycetaceae; Streptomyces.  
OX NCBI\_TaxId=1902;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=A3(2) / M145;  
RX MEDLINE=120096410; PubMed=12000953;  
RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,  
RA Thompson N.R., James K.D., Harris D.E., Quail M.A., Kesser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Frazer A.J., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kesser T., Larre L., Murphy L., Oliver K., O'Neill S.,  
RA Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S.,  
RA Seeger K., Saunders D., Shapard R., Squares S., Taylor K.,  
RA Warren T., Wierzoxrek A., Woodward J., Barrrell B.G., Parkhill J.,  
RA Hopwood D.A. ;  
RT "Complete genome sequence of the model actinomycete Streptomyces  
RT coelicolor A3(2)." ;  
RL Nature 417:141-147(2002) .  
DR EMBL; AL393123; CAC004493.1; -.  
KW Complete proteome.  
SQ SEQUENCE 655 AA; 67324 MW; 523242CD3B7F109 CRC64;

```

RESULT 8
C9FLJ6
ID C9FLJ6 PRELIMINARY; PRT; 674 AA.
AC C9FLJ6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Putative HRP0.
GN PUTATIVE_HRP0.
OS Acidovorax avenae subsp. avenae.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Comamonadaceae; Acidovorax.
CX NCBI_TaxID=80870;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FAVEL8101;
RA Takikawa Y., Eo S., Adachi S., Kojima M.;
RT "Reduced hrp0 genes from Burkholderia glumae and Acidovorax avenae

```

```

      subsp. avenue."')
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL AB053454; BAB20912.1; -
DR GO; GO:0016020; Cmembrane; IEA.
DR GO; GO:0005306; P:protein secretion; IEA.
DR InterPro; IPR001712; Bact_FH1EP.
DR InterPro; IPR006302; HrcV.
DR Pfam; PF00771; FH1EP; 1.
DR PRINTS; PR00949; TYPE3IMAPROT.
DR TIGERfam; TIGR01399; hrcV; 1.
DR Prosite; PS00984; FH1EP; 1.
SQ
SEQUENCE 674 AA; 73214 MW; 73E4B86ED21CBA3C CRC64;

Query Match      2.7%; Score 99.5; DB 2; Length 674;
Best local Similarity 28.5%; Pred. No. 0.15;
Matches 45; Conservative 9; Mismatches 61; Indels 43; Gaps 7.

      226  QYERNRKXKHLVYVVGXG-----ANGAKKXSVGLXNLNKSRSRP----- 265
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db    275  QLTNRITRSYTLAAGLTLPALIPGFPAVPFCFLAAGL----VGGCGMLRSRSRRPPRAAQ 330
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

      226  QY-----XGTLIQRRGASVGLGXDPXPR-----LSPD--AGRPPTSTKXXZAGGVRPPR 309
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db    331  DEFOAHRILPPAARRRQGR--GHGHFAPAPARIROAAVPPVRSAPFPAQCCAAAPFGEEQGR 388
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

      310  AFGPGSXACPSWETGSGRGKGGKGLPLRHAHFHRAAE 347
      |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db    389  RAADAAGAAIPRRRDVGLGRPPG--PALRHPIHQORAK 425

```

RESULT 9	
Q91426	
ID	Q91426
AC	Q91426;
DT	01-OCT-2000 (TReMBLrel. 15, Created)
DT	01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT	01-OCT-2002 (TReMBLrel. 22, Last annotation update)
DE	Gsa protein.
GN	GSA.
OS	Mycobacterium avium.
OC	Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC	Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
OX	NBI_TaxID=1764;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	KEDLINE=20569498; PubMed=1118730;
RA	Bull T.J., Sheridan J.M., Martin H., Sumar N., Tizard M.,
RA	Hermon-Taylor J.;
RT	"Further studies on the GS element: A novel mycobacterial insertion
RT	sequence (TS1612), inserted into an acetylase gene (mpa) in
RT	Mycobacterium avium subsp. silvaticum but not in Mycobacterium avium
RT	subsp. paratuberculosis."
RL	Vet. Microbiol. 77:453-463(2000).
RL	EMBL; A025181; C089383.1; --
Q0	SEQUENCE 377 AA; 41141 MW; BCC5CD7A7A743880 CRC64;



```

Db      239 GYMPFKRKKFLGATYAAHQATFFGASIVAKLGGYDIDFGLIADQLFYRAALIRBSVTID 298
QY      170 VAFKSPDV---GSSYHKEAF---TKRWIVH---PLIG-XXSWVTIVRQ 209
Db      299 RVCVCDPVGPGSGTQPIREHYRLRLMLDHDGYPLGGRVSWAYLRVRE 348

RESULT 10
Q9SDP5 PRELIMINARY; PRT; 750 AA.
ID O9SDP5;
AC O9SDP5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Hypothetical protein.
OS Oryza sativa (rice).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Ehrhartoideae; Oryzaceae; Oryza.
OX NCBI_TaxID=4530;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Nipponbare;
RA Sasaki T., Matsumoto T., Yamamoto K.;
RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
RT clone P0038P12."
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF008036; BAA88189.1; -.
DR Gramene; Q9SDP5; -.
KW Hypothetical protein.
SQ SEQUENCE 750 AA; 82191 MW; C96A063FF1FE2E83 CRC64;

Query Match      2.6%; Score 97.5; DB 10; Length 750;
Best Local Similarity 27.3%; Pred. No. 0.27;
Matches 35; Conservative 16; Mismatches 60; Indels 17; Gaps 5;

QY      236 LVYVLGXGANGAKXXSVGLXNLNASKESRPXGTI---RQRGASVGLGXPRLSPPAGR 292
Db      59 LVVVLGGGSGSGWEERAG--SMEAGPNSAAKKTIMPPRSRGVADRSIFPSRHPPRQ 117
QY      293 PPSTXXXRAGCGVPPRAPGPGSGXACPSWETGRGRKXGKPLARAPHVRAAEYXXS 352
Db      118 PPPPPASRQLAAR-HRQAPP-----PAS-----HRQAPPPIRRSGAALASGHSA 164
QY      353 TTNRRATS 360
Db      165 TVHRRNST 172

RESULT 11
Q8VHD8 PRELIMINARY; PRT; 2496 AA.
ID O8VHD8;
AC O8VHD8;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hornerin.
GN HORNERIN OR 110033K19IRK.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR;
RA MDLINE=21601613; PubMed=11572870;
RA Makino T., Takashi M., Morohashi M., Huh N.-H.;
RT "Hornerin, a Novel Profilaggrin-like Protein and Differentiation-
RT Specific Marker Isolated from Mouse Skin."
RL J. Biol. Chem. 276:47445-47452(2001).
CC -1- SIMILARITY: BELONGS TO THE S-100 FAMILY.
DR EMBL; AY027660; AAK15791.1; -.

```

```

DR      MGD; MG1:1915973; Hornerin.
DR      GO; GO:0001533; C:cornified envelope; IDA.
DR      InterPro; IPR001751; CAP_S100.
DR      InterPro; IPR002048; EF-hand.
DR      Pfam; PF00036; ehand; 1.
DR      Pfam; PF01023; S_100; 1.
DR      ProDom; PD003407; CAP_S100; 1.
DR      ProSite; PS00018; EF_HAND; 1.
DR      ProSite; PS00303; S100_CAP; 1.
DR      PROSITE; PS00303; S100_CAP; 1.
SQ      SEQUENCE 2496 AA; 247588 MW; 4CE136CA6CE657DE CRC64;

Query Match      2.6%; Score 97.5; DB 11; Length 2496;
Best Local Similarity 21.5%; Pred. No. 1.2;
Matches 65; Conservative 30; Mismatches 145; Indels 63; Gaps 12;

QY      58 GEPGALAXGAKRPAARPGATRSXGAPRWVXIGRYT-CQTXQVSGELRE-DRLPLW 115
Db      723 GSTGGQTASSTRQS---GGQASGSGRCGASGQTSGSGSTRYGGSGSRNST 778
QY      116 XRAKARLLIFSTNTXESGASRFPFGFAXXVAKVTGTGLMRPSXSDV-AFXS 174
Db      779 QSRGR-----STRSSTSR---FSGSGSGSGFSGGSGGSGSGSGSGSFG 826
QY      175 FDVGSYHKEAFETKRWIVHPLIGXXSWVTIVRQVAFLLXMCCHGNPAQYERN-RX 233
Db      827 QTEGSGQHGS-----CC-QGSGGYQNEYGS 851
QY      234 RHLYVLGXGANGAKXXSVGLXNLNASKESRPXGTIRQRGASVGLGXPRLSPPAGR 293
Db      852 GHSASGQGQSHYQSSSYGTHNSGSPSPSPAGHSGKSGSLG---QYSGPGQT 907
QY      294 PPSTXXXRAGRV---RRAFPQSGXACPSWETGRGRKXGKPLARAPHVRAAEYXX 350
Db      908 SSTRQSGGQQAASGSGRYGASGQTSGCGSGSTRYGEQSGS-RNSTGSGRSGSTRS 966
QY      351 SST 353
Db      967 SST 969

RESULT 12
Q84357 PRELIMINARY; PRT; 542 AA.
ID Q84357;
AC Q84357;
DT 01-AUG-1998 (TREMBlrel. 07, Created)
DT 01-AUG-1998 (TREMBlrel. 07, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE E2 protein.
GN E2.
OS Mastomys natalensis papillomavirus (Mnpv).
OC Viruses; dsDNA viruses, no RNA stage; Papillomaviridae;
OC Papillomavirus.
OX NCBI_TaxID=10567;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=94120727; PubMed=8291235;
RA Tan C.H., Tachezy R., Van Ranst M., Chan S.Y., Bernard H.U.;
RA Burk R.D.;
RT "The Mastomys natalensis papillomavirus: nucleotide sequence, genome
RT organization, and phylogenetic relationship of a rodent papillomavirus
RT involved in tumorigenesis of cutaneous epithelia."
RL Virology 198:534-541(1994).
RN [2]
RP SEQUENCE FROM N.A.
RA Chan S.;
RA Submitted (SEP-1993) to the EMBL/GenBank/DBJ databases.
RL EMBL; U01834; AA67147.1; -.
DR      HSBP; P03122; 2BP.
DR      GO; GO:0043205; C:host cell nucleus; IEA.
DR      GO; GO:0003700; F:transcription factor activity; IEA.
DR      GO; GO:0006275; P:regulation of DNA replication; IEA.
DR      GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR      InterPro; IPR000427; E2_C.

```

DR InterPro: IPR001866; E2\_N.  
 DR InterPro: IPR009021; Viral\_DNA\_bd.  
 DR Pfam: PF00511; E2\_C; 1.  
 DR Pfam: PF00508; E2\_N; 1.  
 DR ProDom: PD000672; E2\_C; 1.  
 DR ProDom: PD000678; E2\_N; 1.  
 SQ SEQUENCE 542 AA; 60991 MW; E7BB15DF005B26B CRC64;

Query Match 2.6%; Score 96.5; DB 12; Length 542;  
 Best Local Similarity 25.0%; Pred. No. 0.24; Indels 63; Gaps 11;  
 Matches 56; Conservative 16; Mismatches 89;

QY 160 GLW-RPSXHSVAFKSFVYSS-----YHXEA---FTKRWIVHGLXGKSWVTVVR- 208  
 DB 144 GMSRTTSHDINGIFNKGSGDEYVYFKEAKRYSLGTWEVHDLTHSLLIPTVSS 203  
 QY 209 --QVSTFLXMCCHGNPAQYERNRKRLVYVLGKANGAKKXSVGLXNKSSES- 264  
 DB 204 TPQTGFPP-----RQDPVRLHGN-----TTTGLPIPRNSSNOI 237  
 QY 265 -----PXGTRPGRGASVGLGXPR-LSPPAGRPPTTRXXRAGRVPRAPG-- 312  
 DB 238 LAREGKDYDQARRRFRYYQG-PPTPRSLSPPIYRPPPYEERRRKRLRRQDGRV 296  
 QY 313 --FGSXAECPSWET-----GRGRKGXPLARHAPHYRARAE 347  
 DB 297 KYAPSPYRTKPPGETSDDEDEGRGHEPRFQRLPRGLRDGE 340

RESULT 13  
 Q9RSH9 PRELIMINARY; PRT; 552 AA.

AC Q9RSH9; 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TREMblrel. 25, Last annotation update)  
 DE ABC transporter, ATP-binding protein.  
 GN DR2145.  
 OS Deinococcus radiodurans.  
 OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;  
 OC Deinococcaceae; Deinococcus.  
 CX NCBI\_TaxID=1299;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RI / ATCC 13939 / DSM 20539 / NCIB 9279;  
 RX MEDLINE=20036896; PubMed=10567286;  
 RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
 RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
 RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
 RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,  
 RA Makarewicz K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
 RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,  
 RA Fraser C.M.;  
 RT "Genome sequence of the radioresistant bacterium Deinococcus  
 radiodurans R1.";  
 RL Science 286:1571-1577(1999).  
 DR EMBL: AB002048; AAF1688.1; -.  
 DR PIR: F75311; F75311.  
 DR TIGR: DR2145; -.  
 DR GO: GO:0016020; C:membrane; IEA.  
 DR GO: GO:0005524; F:ATP binding; IEA.  
 DR GO: GO:0004009; F:ATP-binding cassette (ABC) transporter acti. .; IEA.  
 DR GO: GO:0000166; F:nucleotide binding; IEA.  
 DR GO: GO:0006810; P:transport; IEA.  
 DR InterPro: IPR003593; AAA\_ATPase.  
 DR InterPro: IPR003439; ABC\_transporter.  
 DR Pfam: PF00005; ABC\_tran; 1.  
 DR SMART: SM00382; AAA; 1.  
 DR PROSITE: PS00211; ABC\_TRANSPORTER\_1; 1.  
 DR PROSITE: PS00893; ABC\_TRANSPORTER\_2; 1.  
 KW ATP-binding; Complete proteome.  
 SQ SEQUENCE 552 AA; 58509 MW; 0688677C8C51C471 CRC64;

Query Match 2.6%; Score 96; DB 16; Length 552;  
 Best Local Similarity 32.1%; Pred. No. 0.27;  
 Matches 42; Conservative 7; Mismatches 44; Indels 38; Gaps 9;

QY 228 ERNRKRLVYVLGKANGAKKXSVGLXNKSSESFPXGTRIRORGAVG--LGXPXPR 285  
 DB 173 DRARRA-----AGHGRTPAVGSDAATTTPVGR--GDYDRGALAGARLSHPLAR 223  
 QY 286 LSPF--AGRPPTTRXX-----RAGRVPR--APPGSXAECPSWETGRGR 329  
 DB 224 AAPFGDAG-PRPDARAPAAARRIRHWPERSGPRRPRRAAPG-----C-----GQHG 272  
 QY 330 KGGXPLARHAP 340  
 DB 273 GAGDPFARGSP 283

RESULT 14  
 Q9HXB4 PRELIMINARY; PRT; 160 AA.

AC Q9HXB4; 01-MAR-2003 (TREMblrel. 23, Created)  
 DT 01-MAR-2003 (TREMblrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)  
 DE Hypothetical protein.  
 OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;  
 OC Cercopithecinae; Macaca.  
 CX NCBI\_TaxID=9541;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Frontal cortex;  
 RA Hashimoto K., Osada N., Hida M., Kusuda J., Sugano S.;  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Frontal cortex;  
 RX MEDLINE=21458551; PubMed=11574149;  
 RA Osada N., Hida M., Kusuda J., Tanuma R., Iseki K., Hirata M., Suto Y.,  
 RA Hirai M., Terao K., Suuki Y., Sugano S., Hashimoto K.;  
 RT "Assignment of 118 novel cDNAs of cynomolgus monkey brain to human  
 chromosomes.";  
 RL Gene 275:31-37(2001).  
 DR EMBL: AB097540; BAC41765.1; -.  
 KW Hypothetical protein.  
 SQ SEQUENCE 160 AA; 16645 MW; 21AAGF74156CAFAC CRC64;

Query Match 2.5%; Score 95.5; DB 6; Length 160;  
 Best Local Similarity 27.0%; Pred. No. 0.069;  
 Matches 44; Conservative 11; Mismatches 43; Indels 65; Gaps 8;

QY 214 LIXMCCGNGNPAQYERNRKRLVYVLGKGA-----NGAKKXSVGLXNKSSES 263  
 DB 8 LILHL-----SAKQGRE-----IGVAGGARSSALQGGGSGSTARSSESAPSRAGS 58  
 QY 264 RFXGTRPGRGASVGLGXPR--RLSPAPARPPS-----TXXPARGRVPR----- 308  
 DB 59 EAPGA--AGRGASVPGLPQCTWBPAPPSRPPPRRVVCRVRRTHRGSGVCAALPAFR 116  
 QY 309 -----RAPGGSXAECPSWETGRKGGXP 334  
 DB 117 SLAGSLCKPSGVGTAGBAAGGS-----RRSGRP 147

RESULT 15  
 Q9XHX5 PRELIMINARY; PRT; 936 AA.

AC Q9XHX5; 01-NOV-1999 (TREMblrel. 12, Created)  
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
 DT 01-OCT-2002 (TREMblrel. 22, Last annotation update)  
 DE Putative polypeptide.

GN OSJNB40049B20.18.  
 OS *Oryza sativa* (Rice).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzoae; *Oryza*.  
 OX NCBI\_taxid=4530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=cv. Nipponbare;  
 RA Bhelli R., Benito M.-I., Lin X., Mason T.M., Umayam L., Shea T.P.,  
 RT Fujii C.Y., Shen M., Fraser C.M.,  
 RL Submitted (IDEC-1999) to the EMBL/GenBank/DBU databases.  
 DR EMBL; AC007789; AAD38291.1; -.  
 DR Gramene; Q9XHX5; -.  
 DR InterPro; IPRO05162; Retrotrans gag.  
 DR Pfam; PF03732; Retrotrans\_gag; 1.  
 KW Polypeptid.  
 SQ SEQUENCE 936 AA; 102715 MW; 374D7D376F3F5FE7 C664;

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

# OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 1, 2004, 15:48:11 ; Search time 27 Seconds  
(without alignments)  
8921.733 Million cell updates/sec

Title: US-10-028-952A-4  
Perfect score: 4230  
Sequence: 1 gcacgagatccactgtcc.....agcttgcactcttcgcca 2333

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 389414 seqs, 51625971 residues  
Total number of hits satisfying chosen parameters: 778828

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ n2p.model -DEV=xlp  
-O=/cgn2\_1/USPTO.spool.p/US10028952.runal.01042004.154626.682/app.query.fasta\_1.2503  
-db=Issued Patents\_AA -Qfmt=fasta -SUPP=raai -MINMATCH=0.1 -LOOBL=0  
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bl0sum62 -TRANS-human40.cdd  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NORM=ext -HEAPSIZE=500 -MILEN=0 -MAXLEN=200000000  
-USER=US10028952 @CGN 1 1 27 @runat.01042004.154626.682 -NCPU=6 -ICPU=3  
-NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS.COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARY

Result No.	Score	Query Match	Length	DB ID	Description
1	1189.5	28.1	355	4	US-08-630-915A-192
2	1033	24.4	538	4	Sequence 192, App
3	1033	24.4	538	4	Sequence 2, Appl
4	702.5	16.6	433	4	Sequence 18, Appl
5	324	7.7	58	4	Sequence 118, App
6	324	7.7	58	4	Sequence 214, App
7	277	6.5	486	2	Sequence 3, Appl
8	261.5	6.2	486	2	Sequence 2, Appl
9	261.5	6.2	486	2	Sequence 26, Appl
10	237	5.6	726	4	Sequence 20675, A
11	236.5	5.6	1418	4	Sequence 32367, A
12	231	5.5	546	2	Sequence 4, Appl

13	231	5.5	546	4	US-08-630-915A-14	Sequence 14, Appl
14	228.5	5.4	798	4	US-09-252-991A-23774	Sequence 23774, A
15	224.5	5.3	663	4	US-09-252-991A-30843	Sequence 30843, A
16	223.5	5.3	663	4	US-09-252-991A-30843	Sequence 19446, A
17	220.5	5.2	957	4	US-09-252-991A-20408	Sequence 20408, A
18	220	5.2	1706	4	US-09-252-991A-31760	Sequence 31760, A
19	219.5	5.2	2294	4	US-09-252-991A-17231	Sequence 17231, A
20	219	5.2	686	4	US-09-252-991A-20509	Sequence 20509, A
21	214	5.1	863	4	US-09-252-991A-26099	Sequence 26099, A
22	213	5.0	1002	4	US-09-252-991A-27980	Sequence 27980, A
23	212	5.0	1008	4	US-09-252-991A-29419	Sequence 29419, A
24	209.5	4.9	720	4	US-09-252-991A-31915	Sequence 31915, A
25	208.5	4.9	783	4	US-09-252-991A-28327	Sequence 28327, A
26	206	4.8	793	4	US-09-252-991A-18035	Sequence 18035, A
27	204.5	4.8	57	4	US-08-630-915A-116	Sequence 116, App
28	204	4.8	594	4	US-09-252-991A-32578	Sequence 32578, A
29	203.5	4.8	555	4	US-09-252-991A-28848	Sequence 28848, A
30	203.5	4.8	891	4	US-09-252-991A-30624	Sequence 30624, A
31	203.5	4.8	1197	4	US-09-252-991A-30833	Sequence 30833, A
32	203	4.8	695	4	US-09-252-991A-29910	Sequence 22910, A
33	202	4.8	654	4	US-09-252-991A-25801	Sequence 25801, A
34	201.5	4.8	1411	4	US-09-252-991A-18008	Sequence 18008, A
35	201	4.8	433	4	US-09-252-991A-27024	Sequence 27024, A
36	199.5	4.7	537	4	US-09-252-991A-20509	Sequence 20509, A
37	199.5	4.7	802	4	US-09-252-991A-25050	Sequence 25050, A
38	199.5	4.7	802	4	US-09-252-991A-17817	Sequence 17817, A
39	199	4.7	472	4	US-09-252-991A-30690	Sequence 30690, A
40	199	4.7	588	4	US-09-252-991A-30690	Sequence 29419, A
41	199	4.7	1008	4	US-09-252-991A-29419	Sequence 37, Appl
42	198.5	4.7	53	2	US-08-942-423-37	Sequence 18557, A
43	198	4.7	679	4	US-09-252-991A-18857	Sequence 31954, A
44	198	4.7	703	4	US-09-252-991A-31954	Sequence 21881, A
45	198	4.7	720	4	US-09-252-991A-21881	

## ALIGNMENTS

RESULT 1  
US-08-630-915A-192  
Sequence 192, Application US/08630915A  
Patent No. 6309820  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: HOFFMAN, No. 6309820h  
APPLICANT: KAY, Brian K.  
APPLICANT: FOWKES, Dana M.  
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
NUMBER OF SEQUENCES: 227  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edwards LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/630,915A  
FILING DATE: 03-APR-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-174  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-5090

TELEFAX: (212) 869-8864/9741  
 TELETYPE: 66141 PENNIE  
 INFORMATION FOR SEQ ID NO: 192:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 355 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: unknown  
 MOLECULE TYPE: peptide  
 US-08-630-915A-192

## Alignment Scores:

Pred. No.:	1,98e-82	Length:	355
Score:	1189.50	Matches:	251
Percent Similarity:	73.28%	Conservative:	15
Best Local Similarity:	69.15%	Mismatches:	63
Query Match:	28.12%	Indels:	34
		Gaps:	6

US-10-028-952A-4 (1-2333) x US-08-630-915A-192 (1-355)

```

QY 724 CTCAGTACGAGAGGAAACCGAGCTTCAGACATTTGGTGTATG-----TCGTTG 771
DB 12 LeuProGlyArgGlyThrProGlyProSerGlyLeuGlyValProGluAspGlnCysArg 31
QY 772 GCTAGAGGACCAATGGGGCGAAGCTACATCTGTGGGATTATGACTGAACGCTCTAAGT 831
DB 32 ValArgAspLeuLysGlyTyrLeuAspSerPheThrAlaLysAlaGluLysGluGlu 51
QY 832 CAGAAATCCCGCCGAGGCGGAAAGATACGGACGCCCGCGAGCTCGCTTGGCTCGAGT 891
DB 52 AsnArgArgLeuGluGluLysArgTyrAlaGluLysAlaGlnArgGlnLeuGluGln 71
QY 892 AGCCGCTCCCCCGCTGTCGCCCGCGGGCGCGCCCGCCCTCCAGCGCCCGCGCGC 951
DB 72 ArgArgGluArgGluLeu-----ArgGluAlaAlaArgArgGluGlnArgTyrGlnGlu 89
QY 952 GCGGAGAGGC---GCGTCCCGCCCGCGCGC----- 979
DB 90 GlnGly-GlyGluAlaSerProGlnSerArgThrTyrGluGlnGlnGluValAla 109
QY 980 -CGGAGACCGGGGTCGGCGCGGAGTCCGCTCGCTGGAAGGAGGCGCGCGGAAA 1038
DB 109 rArgAsnArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 129
QY 1039 GAGCGCCGCGCCCTGCGCCCTGACGACGACGCTTCGTGGGAAACCTGCGCTAAACA 1098
DB 129 LuArg-----AlaMetSer-----ThrT 135
QY 1099 CCTCATCTCCAGTCTCTAGCTGCGAGCTGAGAGGCCCTTCTGAGAGAGAGCTCA 1158
DB 135 hrSerIleSerSerProGlnProGlyLysLeuArgSerProPheLeuGlnLysGlnLeuT 155
QY 1159 CCAACACAGAGACCACTTTGGCAGAGAGGACGAGCTGCTGCATCTCAAGGCCAGGCGAG 1218
DB 155 hrGlnProGluThrhrIlePheGlyArgGluProAlaAlaAlaAlaIleSerArgProArgAla 175
QY 1219 ATCTCCCTGCTAGAGAGCGCGGCCAGACATCTCCCTCATGCTGTGTGAGGCAAGAAGG 1278
DB 175 sPheLeuProAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 195
QY 1279 AGGCTGTATAGAGAACCTCCAGAGAGGAGACCTTCTACAGAGAGGCCCTCACTGTGTC 1338
DB 195 LuAlaValTyrGluGluProProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 215
QY 1339 AGCAGAGAGTGTGCTGTGCTGTGAGCATTGACACCACTTCAGAGAGGCGGCGCTCACTG 1398
DB 215 LngLngLngLngLysArgGlnLysArgGlnLysArgGlnLysArgGlnLysArgGlnLysArg 235
QY 1399 GCGAAGGCTGTGTGCTGTGCTGTGCTGTGACTTACAGAGAGGCGGCGGAGAGAGATCT 1458
DB 235 LysLngLngLngLysArgGlnLysArgGlnLysArgGlnLysArgGlnLysArgGlnLysArg 255

```

```

QY 1459 CCTTGACCCCGAGAACTCATCAAGGAGCATTCAGAGTGTATGAGAGAGGCTGTGCGCTG 1518
DB 255 erPheAspProGluAsnLeuIleThrGlyIleGluValIleAspGlnGlyTyrThrArg 275
QY 1519 GCTATGGGCGGAGTGGCCATTTGGCATTTTCCTCCCACTACGAGAGCTCATTTAGT 1578
DB 275 LysTyrGlyProAspGlyhrIlePheGlyMetPheProAlaAsnTyrValGluLeuIleAsp 294
QY 1579 GAGGCTGAGGACACATCTGTCCTCCCTCCCTGAGACATGAGCTTCTTATTTGTAAGA 1638
DB 295 GluAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 314
QY 1639 GAGGCTGAGGAGTGTGACATTCAGACCTTTCCAGAAATAGAACCCCACTGAGAGATGAG 1698
DB 315 GlyGlyLeuGlyValAlaPheIleGlnIleSerSerArgAsnArgThrProSerGluAspGlu 334
QY 1699 GCGTACGAGCTCCCTCCCGCTGAGAGCTGAGCTGACCCCAATGACAAATGACAAATGAGCC 1758
DB 335 AlaSerGlyLeuProProAlaTyrPheGlnProValThrProAlaAlaMetAla 354
QY 1759 TGG 1761
DB 355 TTP 355

RESULT 2
US-08-994-076-2
; Sequence 2, Application US/08994076
; Patent No. 6500937
; GENERAL INFORMATION:
; APPLICANT: Evid. Jf., Paul R.
; TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCE
; TITLE OF INVENTION: OF MAMMASTATIN AND METHODS OF USE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt
; STREET: 3100 No. 6500937 West Center, 90 South 7th Street
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: Fast-Seq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/994,076
; FILING DATE: 19-DEC-1997
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/027,315
; FILING DATE: 03-OCT-1996
; APPLICATION NUMBER: PCT/US97/16026
; FILING DATE: 03-OCT-1997
; APPLICATION NUMBER: 08/943,828
; FILING DATE: 03-OCT-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ketelberger, Denise M
; REGISTRATION NUMBER: 33,924
; REFERENCE/DOCKET NUMBER: 4273.1US11
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-332-5300
; TELEFAX: 612-332-9081
; TELETYPE:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 538 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
; US-08-994-076-2

```

## Alignment Scores:

Pred. No.: 1,74e-70 Length: 538  
 Score: 1033.00 Matches: 212  
 Percent Similarity: 95.52% Conservative: 1  
 Best Local Similarity: 95.07% Mismatches: 7  
 Query Match: 24.42% Indels: 4  
 DB: 4 Gaps: 0

US-10-028-952a-4 (1-2333) x US-08-994-076-2 (1-538)

QY 417 TTGATCTTGATTTTCAGTACGAATACAGACCGTGAAGCGGGGCTCAGATCCTTCTGA 476  
 Db 1 MetIleuLeuIlePheSerThrAsnThrAspArgGluSerGlyAlaSer-Asp-LeuLeuT 20  
 QY 477 CCTTTGGGTTTAAAGCAGAGAGTGTCAAGAAAGTTACCAAGGATTAAGTGGCTTGTGG 536  
 Db 20 hrPheTrpValLeuSerArgArgCysGlnYsSerTrhSarGAspAsnTrpLeuValA 40  
 QY 537 CGGCAAGCGTTTCATAGCAGCGTGGCTTTTGATCTTCGATGCGGCTCTTCATCAT 596  
 Db 40 laaIalysarGserLysArgArgArgPheLeuIleLeuArgCysArgLeuPheLeuSerL 60  
 QY 597 TGTGAAGCAGAAATTCACCAAGCGTTGATTTGATCCACTAATAGGAAGCTGAGCTGG 656  
 Db 60 euGIysarArgIleHsGlnAlaLeuAspCysSerProThraAsnArgGluArgGluLeuG 80  
 QY 657 GATTAGACCGTGGTGGAGACGTTAGTTTACCTACTGATGATGTGTGTGTCATGCT 716  
 Db 80 lylLeuAspArgArgGlnThrGly-LeuPheTrhLeuLeuMetCysCysGlySgIly 99  
 QY 717 AATCTGCTCAGTACGAGAGAACCGGAGTTTCAGACATTTGTGTATGTCTTGCTGA 776  
 Db 100 AsnProAlaGlnTrpGluArgAsnArgArgPheArgHisLeuValTrpValLeuGlyTrp 119  
 QY 777 GAGGCCAATGGGGCGCAAGCTACCACTGTGGATTTATACAGACGGCTCTTAAGTCAGAA 836  
 Db 120 GlyAlaAsnGlyAlaLysLeuProSerValGlyLeuLeuLysAsnAlaSerLysSerGlu 139  
 QY 837 TCCCGCCGAGGCGGAGCATACGAGCGCGCGGAGCTCGGTTGGCTTGATACCG 896  
 Db 140 SerArgProGlyGlyThrIleArgGlnArgArgGlyAlaSerValGlyLeuGlyTrpPro 159  
 QY 897 GTCCCCCGCTGTCCCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 955  
 Db 160 ValProArgLeuSerProProAlaGlyAlaProProProProAlaProArgAlaArg 179  
 QY 956 GAGGCGCGTGGCGCGCGCGCGCGCGGAGACCGGGGTCCGGTGGGAGTGCCTTCCTCT 1015  
 Db 180 GluGlyAlaCysProAlaAlaArgArgAspArgGlyProValArgSerAlaLeuArgPro 199  
 QY 1016 GGGAAACGGGGCGCGCGCGGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1075  
 Db 200 GlyLysArgGlyAlaAlaGlyLysAlaAlaProSerProValThrHisArgTrhPhe 219  
 QY 1076 GTG 1078  
 Db 220 Val 220

## RESULT 3

US-09-643-476-2

Sequence 2, Application US/09643476

Patent No. 6599495

GENERAL INFORMATION:

APPLICANT: Eryn, Jr., Paul R.

TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCE

OF MAMMASTATTIN AND METHODS OF USE

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSER: Merchant &amp; Gould P.C.

STREET: P.O. Box 2903

CITY: Minneapolis

STATE: MN

COUNTRY: USA

ZIP: 55402-0903

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/643,476

FILING DATE: 22-Aug-2000

CLASSIFICATION: &lt;Unknown&gt;

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/027,315

FILING DATE: 03-OCT-1996

APPLICATION NUMBER: PCT/US97/18026

FILING DATE: 03-OCT-1997

APPLICATION NUMBER: 08/943,828

FILING DATE: 03-OCT-1997

APPLICATION NUMBER: 08/994,076

ATTORNEY/AGENT INFORMATION:

NAME: Weaver, Karie G.

REGISTRATION NUMBER: 43,245

REFERENCE/DOCKET NUMBER: 4273.1USC2

TELECOMMUNICATION INFORMATION:

TELEPHONE: 612-332-5300

TELEFAX: 612-332-9081

TELEX: &lt;Unknown&gt;

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 538 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

FRAGMENT TYPE: internal

SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-643-476-2

US-10-028-952a-4 (1-2333) x US-09-643-476-2 (1-538)

QY 417 TTGATCTTGATTTTCAGTACGAATACAGACCGTGAAGCGGGGCTCAGATCCTTCTGA 476

Db 1 MetIleuLeuIlePheSerThrAsnThrAspArgGluSerGlyAlaSer-Asp-LeuLeuT 20

QY 477 CCTTTGGGTTTAAAGCAGAGAGTGTCAAGAAAGTTACCAAGGATTAAGTGGCTTGTGG 536

Db 20 hrPheTrpValLeuSerArgArgCysGlnYsSerTrhSarGAspAsnTrpLeuValA 40

QY 537 CGGCAAGCGTTTCATAGCAGCGTGGCTTTTGATCTTCGATGCGGCTCTTCATCAT 596

Db 40 laaIalysarGserLysArgArgArgPheLeuIleLeuArgCysArgLeuPheLeuSerL 60

QY 597 TGTGAAGCAGAAATTCACCAAGCGTTGATTTGATCCACTAATAGGAAGCTGAGCTGG 656

Db 60 euGIysarArgIleHsGlnAlaLeuAspCysSerProThraAsnArgGluArgGluLeuG 80

QY 657 GATTAGACCGTGGTGGAGACGTTAGTTTACCTACTGATGATGTGTGTGTCATGCT 716

Db 80 lylLeuAspArgArgGlnThrGly-LeuPheTrhLeuLeuMetCysCysGlySgIly 99

QY 717 AATCTGCTCAGTACGAGAGAACCGGAGTTTCAGACATTTGTGTATGTCTTGCTGA 776

Db 100 AsnProAlaGlnTrpGluArgAsnArgArgPheArgHisLeuValTrpValLeuGlyTrp 119

QY 1076 GTG 1078

Db 220 Val 220

QY	77	GGAGCCAAATGGGGCGAAGTACCACTGTGGGATTTATACATGAAGCGCTCAATGACGAA	836
Db	120	GLYALAAASGLYALALysLeuProSerValGlyLeuLeuAenAlaSerLysSerGlu	139
QY	837	TCCCGCCCAAGCGGAAACGATACGAGCAGCGCGCGAGGCTTCGTGGCTCGGATACCG	856
Db	140	SerArgProdiGlyThrIleArgGlnArgArgGlyAlaSerValGlyLeuGlyTyrPro	159
QY	897	GTCGCCCGCCTGTGCCCGCGCGGCGGGCGCG - CCCCCCTCCACAGCGCGCCGCGCGCGG	955
Db	160	ValProArgLeuSerProAlaGlyAlaProProProProAlaProArgAlaArg	179
QY	956	GAGGGCGCGTGTCCCGCCCGCGCGCGGAGCCGCGGTCCGATGCGAGTGCCTTCGTCT	101
Db	180	GLUGLYALAcysProAlaAlaArgArgAspArgGlyProValArgSerAlaLeuArgPro	199
QY	1016	GGGAAACGGGGCGCGCGCGGAAAGCGCGCGCGCCCTCGCGCTGACGACGACCGCGCTTC	1073
Db	200	GLYArgGlyAlaAlaGlyAlaAlaAlaProSerProValThrIleArgThrPhe	219
QY	1076	GTG 1078	
Db	220	Val 220	
RESULT 4			
US-08-630-915A-18			
Sequence 18, Application US/08630915A			
Patent No. 6309820			
GENERAL INFORMATION:			
APPLICANT: SPARKS, Andrew B.			
APPLICANT: HOPEMAN, No. 6309820h			
APPLICANT: KAY, Brian K.			
APPLICANT: FOWLER, Dana M.			
APPLICANT: MCCONNELL, Stephen J.			
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL			
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND			
TITLE OF INVENTION: USING SAME			
NUMBER OF SEQUENCES: 227			
CORRESPONDENCE ADDRESS:			
ADDRESSEE: Pennie & Edmonds LLP			
STREET: 1155 Avenue of the Americas			
CITY: New York			
STATE: New York			
COUNTRY: USA			
ZIP: 10036-2711			
COMPUTER READABLE FORM:			
MEDIUM TYPE: Floppy disk			
COMPUTER: IBM PC compatible			
OPERATING SYSTEM: PC-DOS/MS-DOS			
SOFTWARE: PatentIn Release #1.0, Version #1.30			
CURRENT APPLICATION DATA:			
APPLICATION NUMBER: US/08/630,915A			
FILING DATE: 03-APR-1996			
CLASSIFICATION: 536			
ATTORNEY/AGENT INFORMATION:			
NAME: Mistrock, S. Leslie			
REGISTRATION NUMBER: 18,872			
REFERENCE/DOCKET NUMBER: 1101-174			
TELECOMMUNICATION INFORMATION:			
TELEPHONE: (212) 790-9090			
TELEFAX: (212) 869-8864/9741			
TELEX: 66141 PENNIE			
INFORMATION FOR SEQ ID NO: 18:			
SEQUENCE CHARACTERISTICS:			
LENGTH: 433 amino acids			
TYPE: amino acid			
STRANDEDNESS:			
TOPOLOGY: unknown			
MOLECULE TYPE: peptide			
US-08-630-915A-18			
Alignment Scores: 2.15e-45 Length: 433			
Pred. No.: 1			

```

Score: 702.50 Matches: 146
Percent Similarity: 72.65% Conservative: 16
Best Local Similarity: 65.47% Mismatches: 35
Query Match: 16.61% Indels: 26
DB: 4 Gaps: 5

US-10-028-952A-4 (1-2333) x US-08-630-915A-18 (1-433)

QY 971 GCCCGCCGCCGCGAC-----CGGGGTCCGGTGGCGAGTGCCCTTCGTGGAA 1021
   |||||:::|||||:::
Db 217 AAlAhAtgAggluGlnArgTyGlIngluInHIsArgSerAlagIAlProSerArg 236
   |||||:::|||||:::
QY 1022 CGGGCGCGCGCGGAAGCGCGCCCCCCTCGCCCGCAGCAACCAGCTGGTG 1081
   |||||:::|||||:::
Db 237 Thrgly-GluPrGluGlnIn-----AlavAlserArghrArgInglutPr 253
   |||||:::|||||:::
QY 1082 AAC-----TGCG 1096
   ::
Db 253 uSerAlagIyGlnGlnAlaProHIsProArgGluIlePheylsgInysgluarAlame 273
   |||||:::|||||:::
QY 1091 CTMAACCCCTCCATCTCCATCTCCAGCTCCAGCGAGAAGTGAAGACCCCTTCAGAA 1155
   |||||:::|||||:::
Db 273 tSerhrThrservAlInrserSerGlnProGlyVysLeurGsrPropheleGlnly 293
   |||||:::|||||:::
QY 1151 GCAGTCAACCCAGCAAGAACCCAATTGGCAGAGAGCCAGCTGTCATCTCAAAGCC 1211
   |||||:::|||||:::
Db 293 sGlnLeuthrGlnProGluInrSertTyGIARgluPrOthrAlaProvalSerArgrpr 313
   |||||:::|||||:::
QY 1211 CAGGCAGATCTCCCTGCTGTAAGAGCGCGGCCCGCACCTCCATGTCAGTCAAGC 1270
   |||||:::|||||:::
Db 313 cAlAlaaglYal---CySGlunGlnProAlaProserThrLeusSer--AlaglnTh 331
   |||||:::|||||:::
QY 1271 AGAAGAGAGGCGTGTGTATGAGAAACTCCAGACAGAGACCTTACGAGACCCCCC 1333
   |||||:::|||||:::
Db 331 rGluGlnGluPrOthrTYrGluValIProPrGluIngsphrIneutyryGluGluPr 351
   |||||:::|||||:::
QY 1331 ACTGTCGACAGAGCAAGGTGGCTGGCTGACACATTGACCAACCATTCAGGGCAGAG 1390
   |||||:::|||||:::
Db 351 clevuValGlnGlnGlnLyAlaglySeGlnHisIleAsparnTyrmelGlnSerGln 371
   |||||:::|||||:::
QY 1391 GCTCAGTGGCGAAGGAGCTCTGTGCCCCGTGCTGTACACTRACGAGCAAGCCAGACAC 1451
   |||||:::|||||:::
Db 371 yPheSerGlyGlnGlyLeuCyAlaArGAlaleutyryspTyrglnAlaalaaSpaph 391
   |||||:::|||||:::
QY 1451 AAAGATTCCTTTGACCCCGAAGACCTATCAAGGCAATCGAGGTGATGACGAAGGCTG 1511
   |||||:::|||||:::
Db 391 rGlnIleSerPheapPrGluInrshneuliethrelYileGluValIIleaSpGlnGlyr 411
   |||||:::|||||:::
QY 1511 GTGGCGTGTCTTAGGCGCGGATGCGCATTTGGCATGTTCCTGCCAATCTAGTGAGGT 1577
   |||||:::|||||:::
Db 411 pTrpaArgGlyTyrglyProaspGlyHIsPhedglyMetPheProAlaentyrValGlu 431
   |||||:::|||||:::
QY 1577 CATTGAG 1577
   |||||:::|||||:::
Db 431 ulleglu 433
   |||||:::|||||:::

RESULT 5
US-08-630-915A-118
; Sequence 118, Application US/08630915A
; Patent No. 6309820
; GENERAL INFORMATION:
; APPLICANT: SPARKS, Andrew B.
; APPLICANT: HOFEMAN, No. 6309820h
; APPLICANT: KAY, Brian K.
; APPLICANT: FOMUKES, Dana M.
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
; TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
; NUMBER OF SEQUENCES: USING SAME
; NUMBER OF SEQUENCES: 227
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Penie & Edmonds LLP
; STREET: 1155 Avenue of the Americas

```



CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/630,915A  
FILING DATE: 03-APR-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Mastrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-174  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIB  
INFORMATION FOR SEQ ID NO: 118:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 58 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-630-915A-118

Alignment Scores:  
Pred. No.: 7,11e-17 Length: 58  
Score: 324.00 Matches: 58  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 7.66% Indels: 0  
Gaps: 0

US-10-028-952a-4 (1-2333) x US-08-630-915A-118 (1-58)

QY 1401 CAAGGGCTGTGTCGCGCCCTGTACGACTACAGGCGAGCGAGACAGAGATCTCC 1460  
DB 1 GIndGlyLeuCyalaArgAlaLeuTyraPtyrGlnAlaAlaAspAspThrGlnLeuSer 20

QY 1461 TTGACCCCGAGAACCTCATCAAGGCGATCGAGTGATGACGAGAGCGTGTGGCGGC 1520  
DB 21 PheAspProGlnuAsnLeuIleThrGlyIleGluValIleAspGluGlyTyrTrpArgGly 40

QY 1521 TATGGCCCGGATGGCCATTTTGGCATGTTCCTGCCCAACTAGCTGAGCTCAT 1574  
DB 41 TyrGlyProAspGlyHisPheGlyMetPheProAlaAsnTyraValGluLeuIle 58

RESULT 6  
US-08-630-915A-214  
Sequence 214, Application US/08630915A  
Patent No. 6309820  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: HOFFMAN, No. 6309820h  
APPLICANT: KAY, Brian K.  
APPLICANT: FOWLER, Dana M.  
APPLICANT: MCCONNELL, Stephen J.  
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
NUMBER OF SEQUENCES: 227  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Penite & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/630,915A  
FILING DATE: 03-APR-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Mastrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-174  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIB  
INFORMATION FOR SEQ ID NO: 214:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 58 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-630-915A-214

Alignment Scores:  
Pred. No.: 7,11e-17 Length: 58  
Score: 324.00 Matches: 58  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 7.66% Indels: 0  
Gaps: 0

US-10-028-952a-4 (1-2333) x US-08-630-915A-214 (1-58)

QY 1401 CAAGGGCTGTGTCGCGCCCTGTACGACTACAGGCGAGCGAGACAGAGATCTCC 1460  
DB 1 GIndGlyLeuCyalaArgAlaLeuTyraPtyrGlnAlaAlaAspAspThrGlnLeuSer 20

QY 1461 TTGACCCCGAGAACCTCATCAAGGCGATCGAGTGATGACGAGAGCGTGTGGCGGC 1520  
DB 21 PheAspProGlnuAsnLeuIleThrGlyIleGluValIleAspGluGlyTyrTrpArgGly 40

QY 1521 TATGGCCCGGATGGCCATTTTGGCATGTTCCTGCCCAACTAGCTGAGCTCAT 1574  
DB 41 TyrGlyProAspGlyHisPheGlyMetPheProAlaAsnTyraValGluLeuIle 58

RESULT 7  
US-08-942-423-3  
Sequence 3, Application US/08942423  
Patent No. 8691673  
GENERAL INFORMATION:  
APPLICANT: Hashimoto, Yasuhiro  
APPLICANT: Takemoto, Yoshihiro  
TITLE OF INVENTION: Lock Binding Protein  
NUMBER OF SEQUENCES: 68  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Syntex (U.S.A.) Inc.  
STREET: 3401 Hillview Ave.  
CITY: Palo Alto  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 94303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/942,423  
FILING DATE: 01-OCT-1997  
CLASSIFICATION: 435

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,715
FILING DATE: 23-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Perles, Rohan
REGISTRATION NUMBER: 35,752
REFERENCE/DOCKET NUMBER: 28260
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 852-1698
TELEFAX: (415) 496-3529
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHEITICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: LCK BINDING PROTEIN
US-08-942-423-3

Alignment Scores:
Pred. No.: 4,81e-13 Length: 486
Score: 277.00 Matches: 87
Percent Similarity: 42.97% Conservative: 23
Best Local Similarity: 33.98% Mismatches: 85
Query Match: 6.55% Indels: 61
DB: Gaps: 2 10

US-10-028-952A-4 (1-2333) x US-08-942-423-3 (1-486)
QY CGCGCCCGCGCGCGCGGAGGCGCGGCGCCCGCGCGCGG-----GACCGGAGT 991
Db 250 Arglysharlguglulgulnlyalaglnlnvalalavagarglulgulnlyarglys 269
QY 992 CCGGTG-----CGAGTGCCTTCCTGTGGAAAGCGGCGCGCGGAAAGCGGCG 1045
Db 270 AlalvalThrlysarYsSerArglulAlaProInProvalIlleAlaMetglulnProla 289
QY 1046 GCCCCCTGCGCCGTCACGACGACCGACGCGTCTGGGGAACCTGGCGCTAAACCATCTCAT 1105
Db 290 ValProAlaProleuProlyllys----- 297
QY 1106 CTCGAGTCTCTAGGCTTGCGAAGCTGAGAGCCCTTCTCTGAGAAAGCAAGCTCACCAACC 1165
Db 298 lIeSerSerGlulAlaTrpProProValglYThrProProSerSergluSergluProval 317
QY 1166 AGAGACCCACTTGGCAGAGACCGACGCTGCATCTCAAGAGCCCGAAGCCAGATCTCC 1225
Db 318 Arg-Thr-----SerArglulhIaProvalProleuLeuProIleArglnThrLeuPr 335
QY 1226 TGCT-----GAGAGCGCGCGCGCCGACGACTCTCTCAAGTGTG-----GTGCGAGC 1270
Db 335 OGluuSpaSnlgulnluPro---ProAlaIleuProIleProArgThrLeuGlulnlglnIva 354
QY 1271 AGAAGAGAGAGGCTGTGTATGAGGAACCTCCAGAGACGAGACCTTCTACGAGCAGCCGCC 1330
Db 354 lGluGlulnluProValTYrglulAlaGlulnProgluProgluPro-----GluProgluPr 372
QY 1331 ACTGTGCAAGCAAGAGTGTGCTGAGCAGCACTTGAACCAACCACTTCAG----- 1382
Db 372 OGluuProgluuaAspTYrgluuSpvaIglulnluMetasparghIsglulglnluuSpgl 392
QY 1382 ----- 1382
Db 392 uProgluIgluAspTYrglulnluValleuGlulnProgluuaPseSerPheSerSerAla 412
QY 1383 -----GGCAGGCGCTCACTGGGCAAGGCGCTCG 1411
Db 412 uAlaIgluSerSergluYCySPeAlaGlYAlaGlYAlaGlYAlaValAlaIleuGlYIleSe 432

```

```

QY      1412  TGCCCCGCGCCCTGTAACATCAACCAAGGAGCCGACGACACACAGAGATCTCTTGAACCCCGA  14711
Db      432   rAlaValAlaIleuNtrAspTyrGlnGlyGluGlySerAspGlnIleuSerPheAspProAs  452
QY      1472  GAAACCTCATCAACGGGCATCGAGGTATCGACGAAAGGCTGGTGGCTGTGGCTATGGGCCCGA  15111
Db      452   PaSpValIleIleThrAspIleGlnIleValAspGlnGlyTyrTyrArgGly--ArgCysHi  471
QY      1532  TGGCCATTGTCGATGTCCTCCCTGGCCCACTACGTCGAACTCATTTAG  1577
Db      471   eGlyHisPheGlyLeuPheProAlaAsnTyrValIleLeuIleuGlu  486

RESULT 8
US-08-942-423-2
Sequence 2, Application US/08942423
Patent No. 5691673
GENERAL INFORMATION:
APPLICANT: Hashimoto, Yasuhiro
APPLICANT: Takemoto, Yoshihiro
TITLE OF INVENTION: Lock Binding Protein
NUMBER OF SEQUENCES: 68
CORRESPONDENCE ADDRESS:
ADDRESSEE: Syntex (U.S.A.) Inc.
STREET: 3401 Hillview Ave.
CITY: Palo Alto
STATE: California
COUNTRY: U.S.A.
ZIP: 94303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/942,423
FILING DATE: 01-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,715
FILING DATE: 23-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Peries, Rohan
REGISTRATION NUMBER: 35,752
REFERENCE/DOCKET NUMBER: 28260
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 852-1698
TELEFAX: (415) 496-3529
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: HSI
US-08-942-423-2

Alignment Scores:
Pred. No.: 7,25e-12 Length: 486
Score: 261.50 Matches: 92
Percent Similarity: 38.26% Conservative: 22
Best Local Similarity: 30.87% Mismatches: 107
Query Match: 6.18% Indels: 77
DB: 2 Gaps: 12

US-10-028-952A-4 (1-2333) x US-08-942-423-2 (1-486)
777 GGAGCAATGGG-----GCGAAGCTACCATCTGTGGATTAATGACTGAACGCTCTTAAG 830

```

330



```

Db      344 -----GluProProlAlaLeuProProArgThrProGluGlyLeuGlnValValGlu 360
Qy      1233 GAGCCGCGCCACAGACATCTCTCCATGTCGTGTCAGGACGACGAAGAGAGAGCGCTGTATGAG 1292
Db      361 GluProValTyrGluValAlaIaIaProGluLeu-----GluProGluProGluProAspTyrGlu 379
Qy      1293 GAACCTCCAGACGACGAGAGACCTTCTACAGACAG----- 1325
Db      380 ProGluProGluThrGluProAspTyrGluAspValGlyGluLeuAlaPArgGlnAspGlu 399
Qy      1326 -----CCCCCATCTGGTCAG 1340
Db      400 AspaIaGluGlyAspTyrGluAspValLeuGluProGluAspThrProSerLeuSerTyr 419
Qy      1341 CAGCAAGGTGCTGGCTCTGAGACATTGACACACACATTCACGGGCGAGGGCTAGTGGG 1400
Db      420 GlnIaGlyProSerAla-----GlyAlaGlyGly 429
Qy      1401 CAAGGAGCTGTGCGCCGCTGACCTGACACTACACAGACGAGCGGACGACACAGATATTC 1460
Db      430 AlaGlyIleSerAlaIleAlaLeuTyrAspTyrGlnGlyGlnGlySerAspGluLeuSer 449
Qy      1461 TTGAACCCCGAAGACCTCATACCGGCGCATCAGGTGATCGACGAAGCGCTGTGCGTGGC 1520
Db      450 PheAspProAspAspIleIleThrAspIleGluMetValaIaPArgGluGlyTyrTPArgGly 469
Qy      1521 TATGGCGCGGATGGCCATTTTGAGCATGTTCCTCGCACTACGAGGAGCTATT 1574
Db      470 Gln---CysArgGlyHisPheGlyLeuPheProAlaSerTyrValValSerLeu 486

```

```

RESULT 10
US-09-252-991A-20675
: Sequence 20675, Application US/09252991A
: Patent No. 6651795
: GENERAL INFORMATION:
: APPLICANT: Marc J. Rubenfield et al.
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
: TITLE OF INVENTION: AERUDINOSA FOR DIAGNOSTICS AND THERAPEUTICS
: FILE REFERENCE: 107196.136
: CURRENT APPLICATION NUMBER: US/09/252,991A
: CURRENT FILING DATE: 1999-02-18
: PRIOR APPLICATION NUMBER: US 60/074,788
: PRIOR FILING DATE: 1998-02-18
: PRIOR APPLICATION NUMBER: US 60/094,190
: PRIOR FILING DATE: 1998-07-27
: NUMBER OF SEQ ID NOS: 33142
: SEQ ID NO 20675
: LENGTH: 726
: TYPE: PRT
: ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-20675

Alignment Scores:
Pred. No.: 5,91e-10 Length: 726
Score: 237.00 Matches: 210
Percent Similarity: 32.21% Conservative: 57
Best Local Similarity: 25.33% Mismatches: 279
Query Match: 5.61% Indels: 283
DB: 4 Gaps: 47

US-10-028-952A-4 (1-2333) x US-09-252-991A-20675 (1-726)
QY 2161 CCCCACTTCCCATAGCCGACGAGGACAGAGAAATGAGAGTGGGGAGCAGACAGCAGG 2102
   ||| ||| ||| ||| :||| :||| :||| :|||
Db 3 ProSerPhePro-----ArgArgSerArgAlaLeuGln-----13
QY 2101 CTCCAACAACGAATTCCTGCTGCTCCCAACACCATGATGCACATTCGATTTGGTCACA 2042
   |||::: |||::: |||:::
Db 14 ---ProLeuArgGlnHisGlyLysThrGlnPro-----23
QY 2041 AACTGTGCTAAATAAACAATTCCTACGTTCACTGTTCCCAAGGTCATTTCTAAACAGTGGT 1982
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24 -----SerAsnArgArgLeuArgLeuArgLeuAlaArgAla-----AspAla 37

```

QY	1991	GGCCACCAACCCCTTAGAA-----GCAAGAGAAAGAAAGAAAGCAAAATTAATGCGAGA	1928
Db	38	GlysnHhSPROLeuSerArgLeuAlaGlyAlaGlnProAlaGlnGlnProGlyGlyArg	57
QY	1927	GGCAAGCAAGACCCCTCCAGATGCCCCCTGCTCAGAACCTTCCCGGACGCTTGGCAGTGGC	1866
Db	58	GlyGlnAlaArgThrArgProAlaProThr-----ThrArgGlyProHis	72
QY	1867	CAGTCAAGGCTTGGACCAAGACAGAGAGGCTTGCTCAGTATCTGTACAGGGCCAAAGCCAA	1808
Db	73	HisProGlyLeu---ArgArgArgValLeuAla-----	82
QY	1867	GCTGTCTGGAGGGTCGGGGGATGCGAGAGATGTGTGGAAATCCAGGCCATTGCTG	1748
Db	83	-----TrpGlnAspArgAlaAspGlnGlnProValLeuArgLeuArgPro-----	98
QY	1747	CATTGGGGTGACAGAGCTGAGTGTGCCAAGCGGAGGGAGCGCTGAGGCTCATCTTCAC	1688
Db	99	-----AlaHisAlaAlaValProArgArgAlaHisHisHisValProHis	113
QY	1687	TGGGGG---TCCTATTCTCGAAGAGTCTCAATGTCACTTCCAGGCTCTCTTCAG	1633
Db	114	ArgAlaValProArgSerAlaArgArgAlaGlnLeuHisProPro-Ala-----Al	130
QY	1630	CAATPAGAACCCATGTCTGAGAGGGAGG------CCAAAGTGTGCC-	1588
Db	130	AspArgSerAspAlaProGlyArgArgGlnArgArgProValGlnAlaArgAlaAlaProLe	150
QY	1587	-----CTCAGCTCACTCAATGAGTCCA-----	1566
Db	150	uGlyGlnHisSerAlaGlnHisArgArgProGlyGlnHisGlyProGlyLeuArgProGln	170
QY	1563	-CGTAGTGGAGGGAGCATGC---CAAAGGCGCATCGGCGCCATGCCACGCCACCGC	1508
Db	170	YArgGlnAspGlnGlyAlaAlaGlyGlyThrGlyAspProAla-----	184
QY	1507	CCTTGTGTGATCACTCGATGCCGCTGATGAGTTCCTGGGGTCMAAGAGATCTCTGTG	1444
Db	185	-----ArgLeuProProGlyCysPro-----GlySerArgArg-----	195
QY	1447	TCGTGCGTGGCTGTGATGCTGTACAGG------CACGGGACAGAGCCCTTGC	1400
Db	196	-ArgAlaArgProGlyAlaArgProGlyAlaValAlaProArgProGlyGlnLeuArgProPr	215
QY	1399	CCACTGACCCCTGSC-----CCTGAATGTGCTGCTCATG	1366
Db	215	AlaAlaAlaLeuProGlyProAlaAspProArgHisAlaThrProGlyCysProAspGlnArg	235
QY	1363	TGCTCAGACAGAGCACTTGTGCTGCGA------CCAGTGGGGCTGTCTCG	1316
Db	235	GAlaGlyAlaAspProValAspAlaAlaGlnArgProGlyAspProValProAlaValAlaX	255
QY	1318	TAGAAG------TCTCTGCTCTGGA	1296
Db	255	GArgArgArgArgCysHisArgLeuGlnHisGlyTyrLeuAlaGlnValaHisProAlaThrAl	275
QY	1297	GATTCTCATATCACAGCCTCTCTTCTGCTGCGACAGAC------	1256
Db	275	ArgGlyArgProGlnGlnProValArgArgAlaGlnGlnAspArgArgAlaLeuGlyArg	295
QY	1257	-----ATGAGAGATGCTGGCGCGCGCTCTTCAGCGGAGAGTCT	121
Db	295	GProPheArgArgAlaLeuArgThrGlnArgHisProAspAspProGlyArgHisGlyPr	315
QY	1216	GCCCTGGGACCTTGAAGATGGACAGACGCTCTCTGCGCAAGATGGGCTCTGGTGGTG	115
Db	315	oPro-----AlaArgProAlaAlaArgArgArgAlaAlaAlaValaGlyGln	330
QY	1156	AGCTGCTCTTCAGAGAGGGGCTCTTCAGAGCTTGCAGAGCTGTGACAGCTGAGATGAGGNG	1097
Db	330	uAlaGlyAlaAlaGlyGlnGly-ProArg-----ArgThrGlyAlaAlaAlaArgT	346

QY		1096 GTTTAGCCCGCAGGTTCCCCACCAAGCTGC-----	1096
Db		346 hrGlInserProGlySerAlaIglSerAlaLeuArgAlaHisArghlaIglUgls	366
QY		1087 --GGTGTGACGGGCGAGGGGGCGGCCGCTTTCCGCGCGGCCCTTTCCCGAGAC	1011
Db		366 erAlaAlaIGlVglUleuArghlaaArgPro----GlyProTrpProAlaIglIngln	384
QY		1009 AAGGCACTCCGACCGACCGGACCCCGGTCCGCGCGCGCGGGGACACGCGCCCTCCCGGCG	950
Db		384 roAlaSerAlaAlaProAlaProGlyGluGly-ArgGlyThrAlaArgAla---AlaArg	402
QY		949 GCGCGGGGCGCTTGAGAGGGGGGGCGGCCCGCGCGGGGAACAGCGCGGGACTAT	890
Db		403 ArgGlnAlaAlaArgGlnArghProThralaValaArgThraSPArghlunspGlnIyArg	422
QY		889 CCGAGGCCAACCGAGGCTCCG-----CGCGCGTCGTCGTCGTCGCG	848
Db		423 ProArgProProProProProProAlaGlnProGluAaspGlnProAlaProAlaAla	442
QY		847 CCTGGGCGGATTCT-----GACTTAAGGGGTTCACTCATAT-----CCCAACAT	800
Db		443 ProGlyArghAspAlaAlaArghAspArghProGlnArghThrlslGlyThrAlaProAlaGlu	462
QY		799 GGTAAGCTTCGCGCCCATGGCTCCTCAAGCCAAGCACAT-----ACACC	758
Db		463 GlyAlaAlaGlnPro---GlnArghlnProAap-HisGlyArghIsGlnProGlyAaspP	481
QY		757 AAATGTCTGAACCTGCGGTCTCTCTCGTACTAGACAG-----GATTACATA	710
Db		481 oGlyAlaLeuProArghGlyGlyAlaArghProArghlnProArghlaArghGlyArghIsGln	501
QY		709 CAACAACACATCATCATAGTAGGTTAAATAACTCTCTCAAGCAGCGTCTAATCCAGCTC	650
Db		501 yGln-----ProHslsGlyArghArghIs-LeuPro----	510
QY		649 ACCTTCCTATTATGAGGGTGAACAATCCAGCGCTTGtGAtTTCGCTTCACATATGATAG	590
Db		510 -----	510
QY		589 GAAGAGCCGACATCGAAGATCAAAAAGCAGCTGCTATGAACGCTTGCCCGCACAAG	530
Db		511 -----ProAlaGln	514
QY		529 CCAAGTTATCCCTGTGTACTTTTGTACACCTCCTGCTTAACCCAAAAAGTCAAGA	470
Db		514 rgGlyGlnPro-----AlaLeuArghProArghAlaAla	526
QY		469 GATCGTAGGCGCCGCTTTCACAGGTC--TGATATCGATAGTAATAATCAAGTCAAGCA	413
Db		526 latnr---GlyProAlaLeuProAlaArghThrValAlaThrgLuglnGlyArg---	543
QY		412 GCTTTGGCCCTTCGTGCTCAAGCGAGTTTTGTGCTCCGTAGCTGCCTTAGGACAC	353
Db		544 -----AlaAlaGluAap	552
QY		352 TGCGTTACGTTTGAACAGGTATACGCCCCAGTCAAATCCCACCTGCACTGTCCCG	293
Db		552 rgAlatyrArghlaThrhProAlaAspProAlaLeuLeuArghIsProGlyAlaGlyG	572
QY		282 GAGCGGGTCCGCGCCCGCGCGCGCGGAGC---CGGCGCTTGAGGCGCAAGCGAGAC	236
Db		572 lyArghProAlaProAlaArghGlyAsnArghlaLeuGlnArghlaArgAlaAap	592
QY		235 CCTCGAGCTCGCCCCCGCGCTCAAGGATGATAAAAAACGATCAAGATAGTGAT	176
Db		592 laAlaAspAlaAlaHisPro-----GlyLyGlnThrVala	604
QY		175 TTCACCGGCGGCGC-----GCAGGGCGCGGACCGCGCC-----	141
Db		605 ----ArglyArghIsAlaProProGlnGlyProGlyProGlyAspProAlaIGlyThrP	623
QY		140 -----CGGGGCCCTTCGCGGGACACCGGGGAGGCGCGGGGCCCAAGCTCACh-	90

```

Db      623 rolaaaglnpfoqglyleuargargvalargarglyl:leaiaaglylthrvlaalaargsera 643
Qy      89 -----GGGTCCTTTCCCGCGCGATTCGCCCA 62
Db      643 rgargh1spropfoProvalAlaProfoleua1aProfoleua1aProglynglycylg 663
Qy      61 AGCCCGTTCCTTGCGT 45
Db      663 lnProAlaargleua1a 668

RESULT 11
US-09-252-991A-32367
; Sequence 32367, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 32367
; LENGTH: 1418
; TYPE: PR
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-32367

Alignment Scores:
Pred. No.: 7,78e-10 Length: 1418
Score: 236.50 Matches: 195
Percent Similarity: 33.02% Conservative: 53
Best Local Similarity: 25.97% Mismatches: 249
Query Match: 5.59% Indels: 255
Db: 4 Gaps: 40

```

```

US-10-028-952A-4 (1-2333) x US-09-252-991A-32367 (1-1418)
Oy 1931 AGAGGAGCCAGCAGGAGCCCTCCAGATGCGCCCTGACAGAGCCCTTCCCC----- 1884
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 10 ArgArgInProAlaArgProLeuArgInGlnHisIaIaIaIaIaProSerAlaAsp 29
    - - - - - GCAGTTTGACAGTGGCCACTCAGAGGCTTGGCCACAGGCGCTTGGCTCA 1833
Oy 1883 - - - - -
    - - - - - GCAGTTTGACAGTGGCCACTCAGAGGCTTGGCCACAGGCGCTTGGCTCA 1833
Db 30 AlaAspPheAlaIaIaValGlnLeuProValArgAlaGlnProAlaGlyPro----- 46
    - - - - - GCAGTTTGACAGGCGCAAGAGCCAGAGCTGTCTGGAGAGGTGCGGCGGATGACAGAGGATG 1773
Oy 1832 GTATCTGTCTCAGGCGCAAGAGCCAGAGCTGTCTGGAGAGGTGCGGCGGATGACAGAGGATG 1773
    - - - - -
Db 46 ----- 46
Oy 1772 TGTTGGAAATCACCAAGGCCATTGCTGCATTGGGGGTACAGGCTAGATGTGCCAAGCCGGA 1713
    - - - - - ArgAlaPro 49
    - - - - -
Db 47 ----- ArgAlaPro 49
Oy 1712 GGGAGCCTCAGAGCCTCATCTCTACGTGGGGTCTCTATTCTGAGAAGTGTGTAATGCA 1653
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 50 GlyAlaProArgThrAlaArgSerLeuGly----- 59
    - - - - -
Oy 1652 ACTCCCAAGGCTCTCTTTCACAGCATAGAAGAGCATGTCTGAGAAGGGAGGAGCCAGAT 1593
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 60 -----ArgProGlnAlaIaIaLeuArgInProGlnHisAspArgLeuGly----- 74
    - - - - -
Oy 1592 GTGCGCTCAGGCTCATCTCAATGAGCTCCAGCTCAGTGTGGCAGGAACATGCCMAATGGCC 1533
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 75 -----HisAlaAspHisArg-HisLeuProAlaThrAl 85
    - - - - -
Oy 1532 ATCCGGCCCATAGCCAGC----- 1514
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```









Fri Apr 2 07:35:03 2004

us-10-028-952a-4.rai

Page 13

Oy 702 ACATCATCAGCAAGGCAATAAAGCTA----- 677  
 Db 376 hvtlaalaaiaabxleuabxghthgaurprogiyaaagabxglaabxglyytrp 396  
 Oy 676 -----TGTCTCAGCAGCGGTAAATCCCAAGCTCAGCTTCCTATTATGGGGTGAACAATCC 622  
 Db 396 rogiyval-Alaagiyabxglyrpheroalaaiaabxghthglaabxglyyysprofo 415  
 Oy 621 AACGCTTGGAATTCCTGCTTCAACATGATAGAAAGCCAGCAT----- 577  
 Db 416 Alaagiy-----LeuAlaaxghthrglyleuabxghthglaabxghthglaabxglyy 433  
 Oy 576 CGAAGGATCAAAAAGCAGCTCGCTAAGAA-----CGCTTGGCCGCGCAAGCGCAG 526  
 Db 434 Argglyalaproaagabxghthglaabxghthgthrproalabxleu-Proglyserylacy 453  
 Oy 525 TTAT-----CCCTGGGTAACCTTTCTGACACCTCCTGCTTAAACC 484  
 Db 453 stlyabxglyalabxserglyrprothnglyalavalaaiaabxprofoalaprofo 473  
 Oy 483 CAAGAAGTCAGAAAGATGTAGAGCCCGCTTTACAGCTGTATTCGTATGAAAATCA 424  
 Db 473 caapabx-----glyrpheroalacysleuabxghthglaabxlaal 486  
 Oy 423 AGATCAAGCAGACTTTGCTCCCTGCTGCTCAGGAGAGGTTCCTCCTCGTAGCTGCG 364  
 Db 486 atyleuabxprothzlaabxetprogiyalaabxprogiyalaabxprofoalabxghthral 506  
 Oy 363 CTTAGACACACTCGGCTTACCGTTTGACAGGTGTACCGCCCA-----GTCAAA 316  
 Db 506 aalaglyylythzabx-Proalabxroabxleuabxprogiyalaabxleuabx 525  
 Oy 315 CTCGCCACCTGACACTGTCGCCCGGAGCGAGTGGCGCCGCGCC----- 274  
 Db 526 leuabxroalabxhbaabxglyhbaabxserglyalabxproalabxghthglaabxleuabx 545  
 Oy 273 -----GGCGCGCGCGCGCGCGCTTGCGCCCGCAAGCGCAGAG 238  
 Db 546 phebghintyrglyalabxghthglaabxghthglaabxghthglaabxghthglaabx 565  
 Oy 238 ----- 238  
 Db 566 leuabxvalabxghthglaabxproabxghthglaabxproabxghthglaabxvala 585  
 Oy 237 -----CCCTCGGGGCTTGCGCCCGCGCGCTCA----- 211  
 Db 586 glyalabxleuabxproalabxghthglaabxghthglaabxghthglaabxproalabx 605  
 Oy 210 -----CCGGGTCAAGTGAATAAAGATCAGAGTAACTGATTTTCA 172  
 Db 606 Argglyalabxghthglaabxghthglaabxghthglaabxghthglaabxghthglaabx 624  
 Oy 171 CGCGAGCGCGCGCGAGCGCGC-----GGACCCCGC 142  
 Db 625 Provalabxghthglaabxghthglaabxghthglaabxghthglaabxghthglaabx 644  
 Oy 141 CCGCGG-----CCCTCGCGCGGAGACCGCGGG 115  
 Db 645 Argthxvalleuabxproleuabxghthglaabxghthglaabxghthglaabxghthglaabx 655  
 RESULT 15  
 US-09-252-991A-30843  
 ; Sequence 30843, Application US/09252991A  
 ; Patent No. 6551795  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Marc J. Rubenfield et al.  
 ; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOWOMANS  
 ; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 ; FILE REFERENCE: 107196.136  
 ; CURRENT APPLICATION NUMBER: US/09/252,991A  
 ; CURRENT FILING DATE: 1999-02-18  
 ; PRIOR APPLICATION NUMBER: US 60/074,788  
 ; PRIOR FILING DATE: 1998-02-18

```

; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 30843
; LENGTH: 663
; TYPE: FRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-30843

Alignment Scores:
Pred. No.:      5,13e-09      Length:      663
Score:          224.50        Matches:      201
Percent Similarity: 30.56%    Conservative: 56
Best Local Similarity: 23.90%  Mismatches:  275
Query Match:     5.31%       Indels:      310
                        Gaps:      47

US-10-028-952A-4 (1-2333) x US-09-252-991A-30843 (1-663)
QY      2192  ATACCCAGGGGGAGGCTTGCCATCTGCATATCCCATCTCCATATAGCCAGGACAGGGA 2133
      1  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      3  ValProAlaProProAlaProAlaTProGlnArgTPro-----ProArgAla 18

QY      2132  CAGAGAAATGAGTGGGGAGCACAGACAGGCTCA-----ACA 2094
      1  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      19  AlalArgAlaSerProTProSerThrThrThrArgProProAlaAspArgSerGlyAlaThr 38

QY      2093  AGCAAAATTCCTCCTGCGCAACACCAACCATGATTCACATCTGACTTGGTCACAAATCTGC 2034
      1  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      39  AlalArgAlaProAlaCysArgArgValProLeArg-----Cys 51

QY      2033  TAAAAACATCTCTACGTTCACTGTTCCCAAGGCGTATTCTAAACAGTGTGGCCACCA 1974
      1  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      52  AlaaSmAlaTProGlyArgProMetSerSerThrSerArgArgProGlyTTrpSerPro 71

QY      1973  CCCCTT-----AGAAGCAAGAGAAAAAGAAAAAGCA 1941
      1  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      72  AlalAlaProAlaGlyAlaCysCysTTrpTrpThrArgSerAlaAlaGlyArgSerAlaThr 91

QY      1940  AATTAATGCAGAGAGCCAGCCAGACCTTCCAGATGCCCCGTGC-----TCAGAGACC 1890
      1  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      92  AspAlaTTrpSerSerAlaProAlaProAlaAsnGlyCysCysLeuPheProAlaGlyPro 111

QY      1889  TTCCCCGACGT-----TGACAGTGGCCACTCAGGCGT-----TGG 1854
      1  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      112  SerProAlaSerProAlaProAlaAlaCysArgArgTrpProArgAlaAlaCysHisTrp 131

QY      1853  CCACAGCAGAGGCTTGCTGCTCAGTATCTCTGTCAGGGGCAAGAGCAAGCTGCTGGAGGG 1794
      1  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      132  ProAlaSer--AlaTTrpTrp-----137

QY      1793  TCGGGGATGCAGAGAAAGATGTGTGGGAATCACCAGGCAATGTGTCATTTGGGGTGACA 1734
      1  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      138  -----TrrPenAlaProAlaArgCysCys-----Trr-----146

QY      1733  GGCTGAGTCTGCCAAGCCGAGAGGAGGCCCTGAGGCTTCATCTCACTGGGGGTCTATTTC 1674
      1  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      147  -----ProAlaP 149

QY      1673  CTGGAAGAGTGTGATATGTCAACTCCAGGCTCTCTTCACAGAAATAGAGAGCCATGT 1614
      1  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      149  ro-----ProAlaProA 153

QY      1613  CTGAGAGGGAGAGGGCAAGATGTGCTCCCTCAGGCTCACTCAATAGAGTCCAGATGTTGGC 1554
      1  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      153  laseAlaGlyAlaAlaCysCysAla--SerProAsnArgArg-----166

QY      1553  AGGAAACATGCCAAATATGCCATCCGGCCCA--TAGCCAGCCCAACGACCTTGCTGCAT 1497
      1  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      167  ---ArgArgGluProTrpProProSerProTyrAlaSerArgAlaGlyProAlaSerCys 185

QY      1496  CACCTCGATGCCCGGTGATGAGGTTCTGGGGTCCAAGAGAAATCTCTGTGTGCTGC-----1442
      1  ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

```

```

Db      185  sGIYAsgProProAlaCysSerProValAlaThrAlaProThrAlaThrCysSerProPr 205
Qy      1441  -----GTCGCTGTAGTTCGTACAGGACGCGGACAGAGCCCTTGC----- 1400
          ::::|
          205  oSerAlaArgSerAlaTrp-----LysProCysAlaCysAl 217
Qy      1399  -----CCACTGAGCCCTGGGCC-----TGAATGTGGTG 1371
          ::::|
          217  alyAlaValAlGlySerAlaArgSerProAlaAsnAlaTrpProAlaAlaSerAlaTrpC 237
          ::::|
          1370  GTCAATGTGCTCAGAGCCAGCACTTGTGCTGCACAGCGGAGGCTGCTGTAGAGAGT 1311
          ::::|
          237  sProThrCysAsnSerAlaArgPro-SerAla-----IleAlaSerThrAlaArgAla 255
          ::::|
          1310  CTCTGCTCTGGAG-----GTTCTCATACACAGCCTCTCTTCTGCTG 1266
          ::::|
          255  expSerThrAsnGlyArgProAlaAlaCysProThrThrProPro----- 271
          ::::|
          1265  CACCAAGATGAGAGAGTGTGCGCGCGGCTCTCAGCAGGAGATCTGCGGAGCT 1206
          ::::|
          272  -----AlaAsnAlaProAlaSerAlaAlaAlaAsnTrpProTrp----- 284
          ::::|
          1205  TGAGATGGCAG-----CAGCTGCTCTCTGCCAAGTGGGTCTGTGTGGGTGAGCTG 1152
          ::::|
          285  -----TrpLysAlaProSerProAlaMetProProSerThrSerAlaThrArg---LeuAla 302
          ::::|
          1151  CTCTGTGAGAGAGGGGCTCCTCAGAGCTGCGAGCTGAGAGCTGAGATGAGAGTGTGTTA 1092
          ::::|
          302  IaAlaCysGlyAlaAspGlyAspAlaCysAlaSerGlnThrProTrpArgAspThrS 322
          ::::|
          1091  GCGGCGAGTTCCCGACCAACGTG-----CGGTCGTGACGGCGGAGGGGCGGCGCG 1041
          ::::|
          322  expProCysValProAsnCysValAsnTrpArgArgGlyLeProTrpSerAlaValAla 342
          ::::|
          1040  CCTTCCGCGCGCGCGCTTCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 983
          ::::|
          342  ystHisSerProTrpProArgTrpProGlyMetProAla-----GlyProArgProS 359
          ::::|
          982  -----CCGCGCGCGCGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 950
          ::::|
          359  erCysThrAlaValAlaAlaAlaTrpAlaProAlaArgGlyGlyTrpAlaAlaAlaProArg 379
          ::::|
          949  -----GCGCGCGCGCGCGTGA 934
          ::::|
          379  snSerCysSerAlaGlyArgLeuArgArgValAlaArgArgSerAlaArgArgAlaTrpAla 399
          ::::|
          933  GG-----GGGGGCGCG-----CCGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 889
          ::::|
          399  rgProTrpProAlaGlyArgAlaThrProAlaSerAlaArgArgAspSer-AlaSer 418
          ::::|
          888  CGAGGCGCAACCGAGGCTCCGGGCGCTGCCGTATCTCCGCTGGCGGAGATTCTGACT 829
          ::::|
          419  ArgGlyArgPro----- 422
          ::::|
          828  TAGAGGCGTTCAGTCATATCCACAGATGTAGCTTCCGCCCAT-----TGAGTC 778
          ::::|
          423  -----AlaThrAlaAlaAspHisProAlaAlaTrpVal 433
          ::::|
          777  CTCAGCCAGACATACCAAAATGTCTGAACCTGCGGTTCC-----TTCGTACT 727
          ::::|
          434  SerAlaAlaArgArgThrSerSerAlaProIleAlaGlySerAlaProGlySerGlyThr 453
          ::::|
          726  GAGCAGAGATTTCAGTGCACACACACATCATCAGTAGGTAAACATACCTGTCTCAGCA 667
          ::::|
          454  AlaProArgCysHis-----ProValArgGlys 462
          ::::|
          666  ---CGGTCTAATCCAGCTCAGTTCCCTTATAGTGGGTGAACAATCAACGCTTGCTGA 610
          ::::|
          463  AspGlyAlaGlyProAlaSer-----ThrGlyArgSerArgArgTrp--- 477
          ::::|
          609  ATTCTGCTTCACATGATGATAGAGAGCGGA----- 580
          ::::|
          478  ---AlaAlaAlaProAlaArgArgAlaArgThrGlyIleArgArgThrAlaAlaAla 496

```

```

Qy      579  -----CATGAGAGATCAAAAAGCAGACGTGCTATGAACG 544
          ::::|
          497  CysTrpProProProAlaProGlyArgArgThrProAlaAlaArgArgSerAlaArg 516
          ::::|
          543  TTGCGCGGACAGACGAGTTATCCCTGTGGTAATTTTCTGACACCTCTCTTAAAC 484
          ::::|
          517  ThrAla-ProArgProAlaIleProAlaHis-----ArgPr 528
          ::::|
          483  CAAGAAGTCA--GAAGAGTGTAGAGCGCCGCTTCAACGCTGTATTCGTACTGAAAA 427
          ::::|
          528  cSerAlaSerAlaThrGlyTrpProGlyProAlaLeuAla----- 541
          ::::|
          426  TCAAGATCAACGAGCTTTTGCCTCTGCTCCACAGGAGGTTCTGCTCCTGAGCT 367
          ::::|
          542  -----CysProAlaAlaAla-GlyArgArgProSer-Pro--Ala 553
          ::::|
          366  CGCCTTAGACACCTGCGTTACCGTTGACAGGTGTAACGCGCAGTCAACCTCCAGC 307
          ::::|
          554  ProAlaProAlaProAla--ProHisAlaArgProAlaArgAlaProArgArgPro-- 571
          ::::|
          306  TGGCACTGCTCCCGAGCGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCTTGCGGCGAG 247
          ::::|
          572  -----GlyArgSerProAlaThrAla--ProAlaAlaAlaAla 584
          ::::|
          246  AAGCAGAGACCCCTGCGGCTGCGCGCGCGCGCGCTCACCGGCTCAGTGAATAACATCAG 187
          ::::|
          585  ProAlaGlyProThrProAlaArgProProAlaAlaAla-----ThrGlyArgPro 601
          ::::|
          186  AGTACTGATATTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 145
          ::::|
          602  SerAlaProProArgProAlaArgArgValArgTrpProTrpArgSerAlaArgCys 621
          ::::|
          144  -----CGCCCGCGCGCGCTGCGCGGAGACACCGCGGAGGCGCGCGCGCGCGCGAGCTC 94
          ::::|
          622  SerProArgArgProThrPro--ArgSerAlaPro-GlyAlaAlaThrGlyAlaProTh 640
          ::::|
          93  AACAGGCTCTTCTTCCCGCTGATTCGCGCAACCGGTTCC 51
          ::::|
          640  rThrGlyAlaAlaArgProArgArgArgAlaAlaAlaPro 654
          ::::|

```

Search completed: April 1, 2004, 15:59:31  
 Job time : 53 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

CM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 1, 2004, 15:47:41 ; Search time 117 seconds  
(without alignments)  
12582.974 Million cell updates/sec

Title: US-10-028-952A-4  
4230  
Perfect score: 1 gcaagagatccaccacgtcc.....agcttggaactcttcgcca 2233  
Sequence:

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Xgapop 10.0 , Xgapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 2034082

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODL=frame+ n2p.model -DEV=xlp  
-O=/cgn2.1/USPTO.spool.p/US10028952/runat.01042004.154625.652/app.query.fasta\_1.2503  
-DB=SPTREMBL\_25 -QWMT=fastan -SUFFIX=resp -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdd -LIST=45  
-DOCALLIGN=200 -THR.SCORE=spct -THR.MAX=100 -THR.MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTMT=Pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US10028952@cgn 1.1.128 @runat.01042004.154625.652 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEOUTERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPTREMBL\_25:\*  
1: sp\_atchea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mmc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvius:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	877	20.7	430	4	Q9NR72	Q9nr72 homo sapien

2	872	20.6	302	4	Q9HBN8	Q9hbn8 homo sapien
3	872	20.6	430	4	Q9UTU6	Q9jut6 homo sapien
4	870	20.6	431	4	Q9CF30	Q9cf30 homo sapien
5	870	20.6	433	4	Q96K74	Q96k74 homo sapien
6	759.5	18.0	282	10	Q9AVH2	Q9avh2 pismu sativ
7	703.5	16.6	436	11	Q80WP1	Q80wp1 mus musculu
8	702.5	16.6	433	11	Q62418	Q62418 mus musculu
9	694.5	16.4	432	11	Q8BH56	Q8bh56 mus musculu
10	665.5	15.7	436	11	Q9JH74	Q9jht4 rattus norv
11	665	15.7	429	11	Q9JH74	Q9jht4 rattus norv
12	661.5	15.6	432	11	Q9JH74	Q9jht4 rattus norv
13	661	15.6	433	11	Q9JH74	Q9jht4 rattus norv
14	392.5	9.3	447	13	Q7ZXQ9	Q7zxq9 xenopus lae
15	288.5	6.8	643	5	Q9XUT0	Q9xut0 caenorhabd
16	265.5	6.3	384	5	Q8KRS9	Q8krs9 dirosophila
17	265.5	6.3	531	5	Q9YU84	Q9yu84 dirosophila
18	261.5	6.2	486	11	Q922T8	Q922t8 mus musculu
19	255	6.0	559	5	Q9YDP4	Q9ydp4 dirosophila
20	255	6.0	587	5	Q96459	Q96459 strongiloe
21	253.5	6.0	559	5	Q96046	Q96046 dirosophila
22	243	5.7	270	10	Q64410	Q64410 zea mays (m
23	236	5.6	471	11	Q70419	Q70419 rattus norv
24	236	5.6	508	11	Q70420	Q70420 rattus norv
25	233	5.5	477	5	Q96620	Q96620 suberites d
26	233	5.5	513	4	Q96H99	Q96h99 homo sapien
27	231	5.5	509	11	Q921L6	Q921l6 mus musculu
28	231	5.5	509	11	Q8BNAS	Q8bnas mus musculu
29	231	5.5	530	13	Q8UWC3	Q8uwc3 xenopus lae
30	220.5	5.2	774	12	Q41971	Q41971 murid herpe
31	213.5	5.0	706	12	Q41972	Q41972 murid herpe
32	213.5	5.0	727	12	Q41973	Q41973 murid herpe
33	209	4.9	890	12	Q8UZE1	Q8uze1 cercopithec
34	206	4.9	206	13	Q7SXH4	Q7sxh4 brachydanto
35	206	4.9	1478	5	Q961U4	Q961u4 dirosophila
36	206	4.9	1843	5	Q9V6W9	Q9v6w9 dirosophila
37	204	4.8	1463	16	Q9ADP6	Q9adp6 streptomyce
38	203.5	4.8	608	12	Q9QSK9	Q9qsk9 herpesvirus
39	201.5	4.8	1111	16	Q9J0P1	Q9j0p1 streptomyce
40	200.5	4.7	758	5	Q86KIT6	Q86kit6 dictyostell
41	199	4.7	530	10	Q04892	Q04892 nictotiana t
42	198.5	4.7	634	4	Q8N707	Q8n707 homo sapien
43	198	4.7	745	16	Q89X06	Q89x06 bradyrhizob
44	196	4.6	579	2	Q86FF2	Q86ff2 streptomyce
45	194.5	4.6	368	3	Q9P837	Q9p837 candida alb

#### ALIGNMENTS

RESULT 1  
Q9NR72 PRELIMINARY; PRT; 430 AA.  
ID Q9NR72  
AC Q9NR72  
DT 01-OCT-2000 (TEMBUREL. 15, Created)  
DT 01-OCT-2000 (TEMBUREL. 15, Last sequence update)  
DT 01-OCT-2003 (TEMBUREL. 25, Last annotation update)  
DE Cervical SH3P7 (Mucin-associated protein).  
GN CMAP.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Ho S.B., Toribara N.W., Anway R.E., Spurr-Michaud S.J., Shekels L.L.,  
RA Keutmann H.T., Hill J.A., Gibson I.K.,  
RT "Expression cloning of human cervical proteins using an antibody to  
RT cervical mucus."  
RT Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
RL  
RU  
RP SEQUENCE FROM N.A.  
RC TISSUE=Cervix;  
RX MEDLINE=99077625; PubMed=9858486;  
RA Gibson I.K., Spurr-Michaud S., Moccia R., Zhan Q., Toribara N.,

RA Ho S.B., Gargiulo A.R., Hill J.A. III,  
RT "MUC4 and MUC5B transcripts are the prevalent mucin messenger  
RL ribonucleic acids of the human endocervix.";  
Biol. Reprod. 60:58-64(1999).  
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
DR EMBL; AF250287; AA81273.1; -.  
DR EMBL; AF151364; AA813120.1; -.  
DR HSSP; P23727; 1PNV.  
DR GO; GO:0005622; C:intracellular; IEA.  
DR GO; GO:0003779; F:actin binding; IEA.  
DR InterPro; IPR002108; Actbind\_cofin.  
DR InterPro; IPR001452; SH3.  
DR Pfam; PF00244; Cofilin\_ADF; 1.  
DR Pfam; PF00018; SH3; 1.  
DR PRINTS; PR00452; SH3DOMAIN.  
DR PRODOM; PD000066; SH3; 1.  
DR SMART; SM00102; ADF; 1.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PS50002; SH3; 1.  
KW SH3 domain. 430 AA; 48154 MW; 700CAP17D8CEB9DC CRC64;  
SQ SEQUENCE 430 AA; 48154 MW; 700CAP17D8CEB9DC CRC64;

Alignment Scores:  
Pred. No.: 1,32e-56 Length: 430  
Score: 877.00 Matches: 180  
Percent Similarity: 74.60% Conservative: 8  
Best Local Similarity: 71.43% Mismatches: 56  
Query Match: 20.73% Indels: 9  
DB: Gaps: 2

US-10-028-952A-4 (1-2333) x Q9NR72 (1-430)

```

QY 841 GCCCAGCGGAGAGAGATACGAGGAGCGCGGAGCGCTGGTGGCTCGATAGCGGCTCC 900
   |||:|||||:|||||:|-----GluGluGluAsnArgArgLeuGluGluValArgArgAla 198
Db 181 AlAlYsAlAGluYs-----GluGluGluAsnArgArgLeuGluGluValArgArgAla 198
QY 901 CCGCGCTGCTCCCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGAGG 960
   |||:|||||:|||||:|-----GluGluGluGluGluGluGluGluGluGluGluGluGlu 218
Db 199 GluGluAlaGluArgGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 218
QY 961 CGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAA 1020
   |||:|||||:|||||:|-----GluGluGluGluGluGluGluGluGluGluGluGluGlu 1020
Db 219 ArgArgGluGluArgGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 238
QY 1021 ACGGG-----GCGCGCGCGGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCTCA 1061
   |||:|||||:|||||:|-----GluGluGluGluGluGluGluGluGluGluGluGluGlu 1121
Db 239 GluGluGluGluValValSerArgAsnArgAsnArgGluGluGluGluGluGluGluGlu 258
QY 1062 CGCAGCGGACGCTCGTGGGAGACCTGGGCGCTAAACACCTCCATCTCCAGTCCAGGCT 1121
   |||:|||||:|||||:|-----GluGluGluGluGluGluGluGluGluGluGluGluGlu 1121
Db 259 GluIlePheIysGluIleGluArgAlaMetSerThrThrSerIleSerSerProGluPro 278
QY 1122 GSCAAGCTGAGAGCGCCCTCTCTCGAGAGAGAGCTCACCAACAGAGAGAGAGAGAGAG 1181
   |||:|||||:|||||:|-----GluGluGluGluGluGluGluGluGluGluGluGluGlu 1181
Db 279 GlyIleValArgSerProPheLeuGluGluGluGluGluGluGluGluGluGluGluGlu 298
QY 1182 AAGAGCGAGCTGCTGCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1241
   |||:|||||:|||||:|-----GluGluGluGluGluGluGluGluGluGluGluGluGlu 1241
Db 299 ArgGluProAlaAlaAlaIleSerArgProArgAlaAspLeuProAlaGluGluProAla 318
QY 1242 CCAGAGCTCTCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1301
   |||:|||||:|||||:|-----GluGluGluGluGluGluGluGluGluGluGluGluGlu 1301
Db 319 ProSerThrProProCysLeuValGluAlaGluGluGluGluGluGluGluGluGluGlu 338
QY 1302 GAGCAGAGAGAGCTTCTAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1361
   |||:|||||:|||||:|-----GluGluGluGluGluGluGluGluGluGluGluGluGlu 1361
Db 339 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 358
QY 1362 CACATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1421
   |||:|||||:|||||:|-----GluGluGluGluGluGluGluGluGluGluGluGluGlu 1421
Db 359 HisIleAspHisIleIleGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 378
QY 1422 CTGTACGACTACGAGGAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1481
   |||:|||||:|||||:|-----GluGluGluGluGluGluGluGluGluGluGluGluGlu 1481

```

Db 379 LeuIleAspArgGluAlaAlaAspHisIleGluIleSerPheAspProGluAsnLeuIle 398

QY 1482 ACGGCGATGAGATGATGAG 1541

Db 399 ThrGlyIleGluValIleAspGluGlyTyrPyrArgGlyTyrGlyProAspGlyHisPhe 418

QY 1542 GGCATGTCCTGCGCAACTAGTGGAGGCTATTGAG 1577

Db 419 GlyMetPheProAlaAsnTyrValGluLeuIleGlu 430

RESULT 2  
Q9HNB8 PRELIMINARY; PRT; 302 AA.

ID Q9HNB8  
AC Q9HNB8; 01-MAR-2001 (TRENDEL. 16, Created)  
DT 01-MAR-2001 (TRENDEL. 16, Last sequence update)  
DT 01-MAR-2003 (TRENDEL. 23, Last annotation update)  
DE Hypothetical protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
OX NCBT\_Taxid=9606;  
RN (1)  
RP SEQUENCE FROM N.A. Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,  
RA Gu J.R., Wan D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,  
RA Qin W.X., Huang Y., Qiu X.X., Qian L.F., He L.P., Li H.N., Yu Y.,  
RA Yu J., Han L.H.;  
RT "Novel Human CDNA clones with function of inhibiting cancer cell  
RT growth." (DEC-1999) to the EMBL/GenBank/DBJ databases.  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
DR EMBL; AF218020; AA817262.1; -.  
DR HSSP; P23727; 1PNV.  
DR InterPro; IPR001452; SH3.  
DR Pfam; PF00018; SH3; 1.  
DR Pfam; PF000452; SH3DOMAIN.  
DR PRINTS; PR000066; SH3; 1.  
DR PRODOM; PD000066; SH3; 1.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PS50002; SH3; 1.  
KW Hypothetical protein; SH3 domain.  
SQ SEQUENCE 302 AA; 34388 MW; 2DECB14C7511612 CRC64;

Alignment Scores:  
Pred. No.: 2,97e-56 Length: 302  
Score: 872.00 Matches: 179  
Percent Similarity: 74.21% Conservative: 8  
Best Local Similarity: 71.03% Mismatches: 57  
Query Match: 20.61% Indels: 9  
DB: Gaps: 2

US-10-028-952A-4 (1-2333) x Q9HNB8 (1-302)

```

QY 841 GCCCAGCGGAGAGATACGAGGAGCGCGGAGCGCTGGTGGCTCGATAGCGGCTCC 900
   |||:|||||:|||||:|-----GluGluGluAsnArgArgLeuGluGluValArgArgAla 70
Db 53 AlAlYsAlAGluYs-----GluGluGluAsnArgArgLeuGluGluValArgArgAla 70
QY 901 CCGCGCTGCTCCCGCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGAGG 960
   |||:|||||:|||||:|-----GluGluGluGluGluGluGluGluGluGluGluGluGlu 50
Db 71 GluGluAlaGluArgGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 50
QY 961 CGCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 1020
   |||:|||||:|||||:|-----GluGluGluGluGluGluGluGluGluGluGluGluGlu 110
Db 91 ArgArgGluGluArgGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 110
QY 1021 ACGGG-----GCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1061
   |||:|||||:|||||:|-----GluGluGluGluGluGluGluGluGluGluGluGluGlu 1061
Db 111 GluGluGluGluValValSerArgAsnArgAsnArgGluGluGluGluGluGluGluGlu 130
QY 1062 CGCAGCGAGCTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1121
   |||:|||||:|||||:|-----GluGluGluGluGluGluGluGluGluGluGluGluGlu 1121
Db 131 GluIlePheIysGluIleGluArgAlaMetSerThrThrSerIleSerSerProGluPro 150

```

QY	1122	GGCAAGCTGAGAGAGCCCTTCCTGTCAGAGAGAGCTCACCAACGAGACCACTTGGC	1181
Db	151	GLYtSLeuAaYgSerProPheLeuclnLysGlnLeuThrGlnProGluThrHisPheGly	170
QY	1182	AGAAAGCCAGTGTCTGCATCTCAAGAGCCAGAGGAGATCTCCCTGCTGAGAGAGCCGGCG	1241
Db	171	ArgGluProAlaAlaAlaAlaIleSerProArgAlaAspLeuProAlaGluInProAla	190
QY	1242	CCGAGACTCTCTCCATGTCTGTGTGAGGAGAGAGAGAGAGAGGCTGTATAGAGAACTTCCA	1301
Db	191	ProSerThrProProCysLeuValGlnIleGluGlnGluAlaValItyGluGluProPro	210
QY	1302	GAGCAGGAGACTTCTTACGAGAGAGCCCACTGGTCGAGAGGAGAGAGTGTGGCTCTAG	1361
Db	211	GIuGlnGluThrPheTyrGluInProProLeuValGlnGlnGlnGlyAlaIleYserGlu	230
QY	1362	CACATTTGACCAACACATTCAGGGCCAGGGGCTCAGTGGGCAAGAGGCTCTGTGCCCTGCC	1421
Db	231	HisIleAspHisHisIleGlnGlyInGlyLeuSerGlyGlnGlyLeuCysAlaArgAla	250
QY	1422	CTGTACGACTACCAAGAGAGCCAGACACAGAGACTCTCTTGGACCCGGAGAGACTCATC	1481
Db	251	LeuTyrAspTyrGlnAlaAlaAspAspThrGluIleSerPheAspProGluAsnLeuIle	270
QY	1482	ACGGGCACTGAGAGTGATCGACCGAAGCTGTGTGCGTGTGCTATAGGCGCGATGCCATTT	1541
Db	271	ThrGlyIleGluValIleAspGluGlyTyrProArgGlyTyrGlyProAspLysHisPhe	290
QY	1542	GGCATGTTCCCTGCCAATCAGCTGAGAGGCTGATGAG	1577
Db	291	GlyMetPheProAlaAsnTyrValGluLeuIleGln	302
RESULT 3			
Q9UJ6			
ID	Q9UJ6	PRELIMINARY;	PRT; 430 AA.
AC	Q9UJ6		
DT	01-MAY-2000	(TREMBlrel. 13, Created)	
DT	01-MAY-2000	(TREMBlrel. 13, Last sequence update)	
DT	01-OCT-2003	(TREMBlrel. 25, Last annotation update)	
DE	SRC homology 3 domain-containing protein H1P-55 (Drebrin F).		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.		
CN	NCBI_TaxID=9606;		
XX	[1]		
RP	SEQUENCE FROM N.A.		
FX	MEDLINE=20036526; PubMed=10567356;		
RA	Euseanal D., Yao Z., Wang X.-S., Kori R., Zhou G., Lee S.C., Tan T.-H.;		
RT	"A Novel Src Homology 3 Domain-containing Adaptor Protein, H1P-55,		
RT	that Interacts with Hematopoietic Progenitor Kinase 1.";		
RL	J. Biol. Chem. 274:33945-33950(1999).		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RA	Zhang W., Yuan Z., Wan T., He L., Cao X.;		
RT	"Molecular cloning of cDNA encoding drebrin F.";		
RL	Submitted (JUL-1998) to the EMBL/Genbank/DBD databases.		
RN	[3]		
RP	SEQUENCE FROM N.A.		
PC	TISSUE=Brain;		
RL	Submitted (JUN-2002) to the EMBL/Genbank/DBD databases.		
DR	EMBL, AF197050; RAFL3701.1; -		
DR	EMBL, AF077353; AAF80228.1; -		
DR	EMBL, BC031687; AAH31687.1; -		
DR	HSSP; P23727; IPNO.		
DR	GO; GO:0008047; Enzyme activator activity; TAS.		
DR	GO; GO:0005515; Protein binding; TAS.		
DR	GO; GO:0007257; Protein binding; TAS.		
DR	InterPro; IPR002108; Acbind_cofln.		
DR	InterPro; IPR001452; SH3.		
DR	Pfam; PF00241; coflin_ADF; 1.		
DR	Pfam; PF00018; SH3_1.		
DR	PRINTS; PR00452; SH3DOMAIN.		

[illegible]

DE Similar to src homolog 3 domain-containing protein HIP-55.  
OS Homo sapiens (Human) .  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_taxid=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE=Eye;  
RA Strausberg R.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ  
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
DR EMBL: BC011677; AAI11677.1; -  
DR GO: GO:0005622; C:intracellular; IEA.  
DR GO: GO:0003775; F:actin binding; IEA.  
DR InterPro: IPR002108; Actbind\_colfin.  
DR InterPro: IPR01452; SH3.  
DR Pfam: PF00241; cofilin\_ADPf\_1.  
DR Pfam: PF00018; SH3\_1.  
DR PRINTS: PR00452; SH3DOMAIN.  
DR ProDom: PD00066; SH3\_1.  
DR SMART: SMO0102; ADP\_1.  
DR SMART: SMO0326; SH3\_1.  
DR PROSITE: PS50002; SH3\_1.  
KW SH3 domain.  
KW SEQUENCE 431 AA; 48294 MW; D18F9D31EFFD0B4E CRC64;

Alignment Scores:			
Pred. No.:	4,38e-56	Length:	431
Score:	870.00	Matches:	161
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	20.5%	Indels:	0
DB:	4	Gaps:	0

QY	1095	ACCACTTCGATCTCCAGTCCCTCAGCTCTGGCAAGCTGAGAGAGCCCTTCTCTGAGAGAGG	115
Db	271	ThlrNrSrlIeSerSerProGlnProGlyLysLueAArgSerProPheNeuGlnLysGln	290
QY	1155	CTCAACCCACAGAGAGACCACCTTTTGAGCAGAGAGAGCCAGCTGCTGCCATCTCAAGGCCCAGG	1214
Db	291	LeuthrGlnProGlnThlrHrIlePheGlyArgGlnProAlaIleAlaIleSerArgProArg	310
QY	1215	GCAAGTCTCCCTGCTGAGAGAGCCGGCGCCAGCACTCCTCCATGTCGTGAGAGCGAGAA	1274
Db	311	AlaPheLueProAlaGlnGlnProAlaProSerTrnProProCysLueValGlnAlaGln	330
QY	1275	GAGAGAGCTGTGTATGAGAGAACCTCCAGAGCAGAGACCTTCTTACAGACAGGCCCCACTG	1333
Db	331	GlnGlnAlaValTyrGlnGlnLueProGlnGlnGlnTrnPheTyrGlnGlnLueProLue	350
QY	1335	GTGACAGCAGCAAGGTGCTGGCTCTGAGCATTGACACACACATTCAGAGGCCAGGGGCTC	1399
Db	351	ValGlnGlnGlnGlyAlaGlySerGlnHsIleAphHsHsIleGlnGlyGlnLue	370
QY	1395	AGTGGGCAAGGGCTGTGTGCTCCGTCGCTCCCTGTACGACTACAGAGCAGCCGACGACACAGAG	1455
Db	371	SerGlyGlnGlyLueCysAlaArgAlaLueTyrIleTyrGlnAlaAlaAspHsPheTrnGln	390
QY	1455	ATCTCTCTTGAACCCCGAGAACTCATACAGGGCATCGAGGTGATGAGAGAGCGGTGG	1511
Db	391	IleSerPheAspProGlnLueTrnLueTrnGlyIleGlyValIleAspLueGlyTrp	410
QY	1515	CGTGCGCTATGGGCGCGATGGCCATTGTGGCATGTTCCCTGCCAACTACGTGAGGTCAAT	1577
Db	411	ArgGlyTyrGlyProAspGlyHrIlePheGlyMetPheProAlaAsnTyrValGlnLueIle	430
QY	1575	GAG 1577	
Db	431	Gln 431	

```

096K74 ID 096K74 PRELIMINARY; PRT; 439 AA.
AC 096K74;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein FLJ14461.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Isgaol T., Oca T., Hayashi K., Sugiyama T., Otsuli T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Watabeuma M., Hosofiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Wakarabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,
RA Nimomiya K., Iwayanagi T.;
RL "NEDO human cDNA sequencing project.";
RT Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AK027367; BAB55065.1; '-'
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003779; F:actin binding; IEA.
DR InterPro; IPR002108; Actbind_cofin.
DR InterPro; IPR001452; SH3
DR Pfam; PF00241; cofillin_ADF_1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000066; SH3; 1.
DR SMART; SMO0102; ADF; 1.
DR SMART; SMO0326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
DR Hypothetical protein; SH3 domain.
KW SEQUENCE 439 AA; 49042 MW; 8C044FBDOE82C6D5 CRC64;

```

Alignment Scores:	
Pred. No.:	4,39e-56
Score:	870.00
Percent Similarity:	100.00%
Best Local Similarity:	100.00%
Query Match:	20.57%
DB:	4
US-10-028-9552A-4 (1-2333) x Q96K74 (1-439)	
Length:	439
Matches:	161
Conservative:	0
Mismatches:	0
Indels:	0
Gaps:	0

QY	1095	LCQKSTCCGATCTCCAGGTCCTCAGCCCTGGAGAGGAGGAGCCCTTCTCTCAGAAAGAG	1157
Db	279	ThrnThrSerLeuSerProGlnProGlyLeuArgSerProPheLeuGlnIysGln	298
QY	1155	CTCAACCCACAGAGAGACCCACTTTGGCAGAGAGCCAGCTGCTGCATCTCAAGGCCACAG	1211
Db	299	LeuThrGlnProGlnThrHisPheGlyValGlnProAlaAlaAlaLeuSerArgProArg	318
QY	1215	GCAGATCTCCCTGGCTGAGAGAGCCGCGGCCAGCACTCCTCATGCTCTGTGCAGAGCAAA	1274
Db	319	AlaAspLeuProAlaGlnGlnProAlaProSerThrProProGlyLeuValGlnAlaGln	338
QY	1275	GAGGAGGCTGTGTATGAGAACTTCGAGAGCAGAGAGACTTCTTACGACAGACCCCACTG	1333
Db	339	GlnGlnAlaValAlaTyrGlnGlnProProGlnGlnGlnThrPheTyrGlnGlnProProLeu	358
QY	1335	GTGCAGACGAGAGGCTGTGCTCTCGAGACACTTGACCACTTCAGGACGACAGGAGCTC	1394
Db	359	ValGlnGlnGlnGlnGlyAlaGlySerGlnHisIleAspHisHisIleGlnGlnGlnIysLeu	378
QY	1395	AGTGGGACAGGAGCTCTGTGCCCCGGTCCCTGTACACACTGACAGAGAGCCGAGAGACAGAG	1455
Db	379	SerGlyGlnGlyLeuCysAlaArgAlaLeuTyrTrpTyrGlnAlaAlaAlaAspSerPheGln	398



```

QY 1455 ATCTCTTGACCCCGAAGAACTTCATACGCGGATCGAGGTGATTCGACGAGGCTGTGG 1514
DB 399 ILeSerPheAspProGluAsnLeuIleThrGlyIleGluValIleAspGluGlyTyrPip 418
QY 1515 CGTGCATGCGCGGATGCGCATTTTGCGATGTCCTCCCACTACGTGAGTCATT 1574
DB 419 ArgGlyTyrClyProAspGlyHisPheGlyMetPheProAlaAsnTyrValGluLeuIle 438
QY 1575 GAG 1577
DB 439 Glu 439

RESULT 6
Q9AVH2
ID Q9AVH2 PRELIMINARY; PRT; 282 AA.
AC Q9AVH2;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE Putative senescence-associated protein (Fragment).
GN SSA-13.
OS Pisum sativum (garden pea).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids I; Fabales; Fabaceae; Papilionoideae; Viciaeae; Pisum.
CX NCBI_Taxid=3888;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Ichihara wase; TISSUE=Pod;
RA Parlaesa U.A.T.; Sunaga A.; Miyazaki T.; Hisaka H.; Sonoda M.,
RA Nakagawa H.; Sato T.;
RT "Cloning of cDNAs encoding senescence-associated genes, ACC synthase
RT and ACC oxidase from stored snow pea pods (Pisum sativum L. var
RT saccharatum) and their expression during pod storage.";
RL Postharvest Biology and Technology 22:239-247(2001).
DR EMBL: AB049723; BAB33421.1; -.
FT NON_TER
SQ SEQUENCE 282 AA; 31822 MW; FC29C7257D2D1059 CRC64;

Alignment Scores:
Pred. No.: 6,52e-48 Length: 282
Score: 759.50 Matches: 187
Percent Similarity: 68.28% Conservative: 11
Best Local Similarity: 64.48% Mismatches: 67
Query Match: 17.96% Indels: 26
DB: Gaps: 5

US-10-028-952a-4 (1-2333) x Q9AVH2 (1-282)
QY 838 GATCTGACTTAAGGGGCTTCAGTCATATCCGACAGATGTAAGTTCGCCCATGGCT 779
DB 11 AspergillusPenicilliumAspergillusAspergillusserpentinusProthioidesPhealProleula 30
QY 778 CCGTGCAGCAACATACATACATATGTCGACCTGCGGCTTCCTCTGTAAGTGAAGA 719
DB 31 PheInProserAlaMetThrAsnCyalaAsnGlnrPheLeuSerTyrTyrValGlu 50
QY 718 TTACATAGGCAACAACATCATATGTAAGGTAAATACTCTGTCACGAGAGGCTCTAA 659
DB 51 Leu--LeuLeuArgHisCysHisGlnTyrPheGlyValThrAsnLeuSerHisAspGlyLeuI 70
QY 658 TCCAGGTCAAGTCCCTATATGAGTGAACAATCCAGCTTGGAATTCGCTTCA 599
DB 70 IeProAlaHisValProTyrTyrPheAlaAsnSerProThrLeuGlyGluPheCysPhe 90
QY 598 CAATGATAGGAAGCCGACATCCAGAGATCAAAAAGCAGCTGCTATGAACGCTTGGC 539
DB 90 hMetIleGlyArgAlaAspIleGluGlySerYssSerAsnValAlaMetAsnAlaTyrP 110
QY 538 CGCCAGACAGCAAGTATCCCTGTGTGTAAGTTCGACACTCTGCTTAAACCAAAA 479
DB 110 eurProGlnAlaSerTyrProCysGlyAsnPheSerPheThrSerPheLys-PheArg 129

```

```

QY 478 GGTCAAGAAAGATCGTAGAGCCCGCTTCAAGGTCTGTATTCTGACTGAATAATCAAGTC 419
DB 130 SerLeuYsAspArgLeuAlaThrLeuSerArgPheValPheValLeuGluIleArgIle 149
QY 418 AAGCAGATTTTGCCCTCTCTGCTCCACGAGAGTTCCTGCTCTGCTGAGCTCGCC-TTA 360
DB 150 LysArgAlaPheThrLeuLeuPheHisThrArgPheLeuPheSerLeuSerSerLeu 169
QY 359 GGACACTGCGCTTACCGTTTACAGAGTTCACCGCCCACTCAACTCCCACTGCGCACT 300
DB 170 GlyHisLeuArgTyrLeuLeuThrAspValProProGlnProAsnSerProAspAsn 189
QY 299 GTCCCGGAGCGGGCGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 249
DB 190 Val-----PheArgPro--AspArgProThrLysValSerLeuThrIle 204
QY 248 AGAAGCGAGAGCCCTCGGGGCTCGCCCGCGCTCAACGGGTACGTGAAAAACGATC 189
DB 204 YsLysGluGlyGln-----ArgArgProProAspSerArgAsnLysLeuAsnValL 222
QY 188 AGAGTAGTGTATTTCACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 129
DB 222 YsSerSerGlyIleSerThrPheAlaValSerSer-----SerHisLeuS 237
QY 128 CGGGGACA-----CCGGGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 81
DB 237 eTyrThrSerGlnValIleSerGlnSerArgThrArgValIleuAsnArgValPhe 256
QY 80 TTCCTCCGCTGATTCGCCCAAGCCGCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCT 21
DB 257 PheProAlaAspSerAlaLysAlaValProLeuLeuValSerLeuAspSerArgYs 276
QY 20 GGACAGTGGCAATCTCGT 3
DB 277 GlyGlnTyrGluPheArg 282

RESULT 7
Q80WPI
ID Q80WPI PRELIMINARY; PRT; 436 AA.
AC Q80WPI;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Actin-binding protein 1.
GN ABP1.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
CX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL6/J;
RA Hou P.; Estrada L.; Goreki J.L.;
RT "Fgbl, a Cdc42-specific guanine nucleotide exchange factor, directly
RT interacts with cortactin and Abp1 to modulate cell shape.";
RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY098595; AAM28340.1; -.
DR GO: GO:0005622; C:intracellular; IEA.
DR GO: GO:0003778; F:actin binding; IEA.
DR InterPro: IPR002108; Actbind_cofin.
DR InterPro: IPR001452; SH3.
DR Pfam: PF00241; coflin_ADF_1.
DR PRINTS: PR00452; SHADOWA1N.
DR ProDom: PDD00066; SH3_1.
DR SMART: SM00102; ADF_1.
DR SMART: SM00326; SH3_1.
DR PROSITE: PS00002; SH3_1.
SQ SEQUENCE 436 AA; 48699 MW; 85AEF9781C698A3F CRC64;

Alignment Scores:
Pred. No.: 9.85e-44 Length: 436
Score: 703.50 Matches: 152

```

Percent Similarity: 66.80% Conservative: 15  
 Best Local Similarity: 60.80% Mismatches: 62  
 Query Match: 16.63% Indels: 22  
 DB: 11 Gaps: 5

US-10-028-952a-4 (1-2333) x Q80WP1 (1-436)

```

QY 871 GAGCCCTGGGTTGGCCCTGAGTACCGCCGCTCCCTGCTCCCGCGGCGGCGCCGCC 930
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 194 GlnuysarArghAlaInuInuArgInuArgInuArgInuArgInuArgInuArgInu 210
QY 931 CCCTCCAGCGCCCGCGCGCGGAGGCGCGCTGCCCCGCCG----- 975
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 211 -----GluArgGluInuInuAlaAlaArgArgInuArgTyArgInuInuHis 228
QY 976 -----GCGCGGAGCGCGGCTCGGAGTCCGCTTCGCTCGGAGAAAG 1023
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 229 ArgSerAlaGlyAlaProSerProSerSerArgInuArgInuProGluInuAlaVal 248
QY 1024 GGGCGCGCGCGGAGGCGCGCGCCCTCGCCCTGAGCAAGCGCGCGTTCGTTGTT 1077
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 249 SerArgThyArgGlnInuInuProGluSerAlaGlnInuAlaProHisProArgGlu 268
QY 1078 -----GGGGAACCTGCGCCCTA-ACCAGCTCCATCTCCAGTCTCCAGCTGCG 1127
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 269 PheLyGlnLyGlnArgAlaMetSerThrThrSerValThrSerSerGlnProGly 288
QY 1128 CTGAGAGAGCCCTCTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1187
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 289 LeuArgSerProPheInuInuInuGlnInuInuProGluInuThySerTyArgInu 308
QY 1188 CCAAGCTGCTCCATCTCAAGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1247
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 309 ProThAlaProValSerArgProAlaAlaGlyVal--CysGlnGluProAlaPro 327
QY 1248 ACTCTCTCAAGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1307
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 328 ThrLeuSerSer--AlaGlnThrGlnGlnGlnProThyArgValProProGluInu 346
QY 1308 GAGACCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1367
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 347 AspThrLeuTyArgInuInuProProLeuValGlnGlnGlnGlnAlaGlySerGln 366
QY 1368 GACACACACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1427
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 367 AspAsnTyArgMetGlnSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 386
QY 1428 GACTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1487
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 387 AspTyGlnAlaAlaAlaAspArgThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 406
QY 1488 ATGAGAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1547
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 407 LLeGlnValAlaLeuArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 426
QY 1548 TTCCTGCAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1577
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 427 PheProAlaAsnTyArgValGlnLeuArgGln 436
  
```

RESULT 8  
 Q62418 PRELIMINARY; PRT; 433 AA.

AC Q62418;  
 DT 01-JUN-1998 (TREMBLrel. 06, Created)  
 DT 01-JUN-1998 (TREMBLrel. 06, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Drebilin-like SH3 domain-containing protein SH3P7.  
 GN Drebilin-like SH3 domain-containing protein SH3P7.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]

SEQUENCE FROM N.A.  
 RP TISSUE=EMBRIO;  
 RC MEDLINE=98294438; PubMed=9630982;  
 RA Sparks A.B., Hoffman N.G., McConnell S.J., Fowlkes D.M., Kay B.K.,  
 RL Nac. Biotechnol. 14:741-744(1996).  
 CC -1- SIMILARITY: CONTAINS A COPY OF THE SH3 DOMAIN.  
 CC -1- SIMILARITY: TO DREBIN.  
 DR EMBL: U58884; AAC52640.1; -.  
 DR HSPSP; P19174; 2HSP.  
 DR MED; MGI:700006; Dbnl.  
 DR GO; GO:0005737; C:cytoplasm; IDA.  
 DR GO; GO:003027; C:lamellipodium; IDA.  
 DR GO; GO:0003779; F:actin binding; IDA.  
 DR GO; GO:0005515; F:protein binding; IDA.  
 DR GO; GO:0016601; P:RAC protein signal transduction; IDA.  
 DR InterPro; IPRO02108; Actbind\_cofin.  
 DR InterPro; IPRO01452; SH3.  
 DR Pfam; PF00241; cofilin\_ADF; 1.  
 DR Pfam; PF00018; SH3; 1.  
 DR PRINTS; PR00452; SH3DOMAIN.  
 DR PRODOM; PD000066; SH3; 1.  
 DR SMART; SM00102; ADF; 1.  
 DR SMART; SM00326; SH3; 1.  
 DR PROSITE; PS50002; SH3; 1.  
 KM SH3 domain. 356 432 SH3.  
 FT DOMAIN 433 AA; 48428 MM; 602D3862C446FA4D CRC64;  
 SQ SEQUENCE 433 AA; 48428 MM; 602D3862C446FA4D CRC64;

Alignment Scores:  
 Pred. No.: 1.17e-43 Length: 433  
 Score: 702.50 Matches: 146  
 Percent Similarity: 72.65% Conservative: 16  
 Best Local Similarity: 65.47% Mismatches: 35  
 Query Match: 16.61% Indels: 26  
 DB: 11 Gaps: 5

US-10-028-952a-4 (1-2333) x Q62418 (1-433)

```

QY 971 GCGCGCGCGCGGAGC-----CGGAGTCGGTGGAGAGTCCCTGCTGGAGAA 1021
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 217 AlaAlaArgArgGlnGlnAlaArgTyArgInuGlnHisArgSerAlaGlyAlaPro 236
QY 1022 CGGCGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1081
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 237 ThrGly-GluProGluInuInu-----AlaValSerArgThyArgGlnInuPro 253
QY 1082 AACC-----TGCGC 1090
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 253 uSerAlaGlyGlnAlaProHisProArgGluLeuPheLyGlnLyGlnArgAlaMet 273
QY 1091 CTAACACACTCCATCTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCCAGTCC 1150
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 273 tSerThrThrSerValThrSerSerGlnProGlyLyLeuArgSerProPheLeuGln 293
QY 1151 GCAGCTACCCAGACCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1210
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 293 sGlnLeuThrGlnProGluInuThySerTyArg-LysGlnProThAlaProValSer 313
QY 1211 CAGGAGAGATCCCTCCCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1270
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 313 oAlaAlaGlyVal--CysGlnGluProAlaProSerThrLeuSerSer--AlaGlnTh 331
QY 1271 AGAAGAGAGAGCTGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1330
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 331 rGlnGlnGlnProThyArgGlnValProProGluGlnAspThrLeuTyArgInuPro 351
QY 1331 ACTGTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1390
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 351 oLeuValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 371
QY 1391 GCTCAGTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1450
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 371 yPheSerGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 391
  
```

QY 1451 AGAATCTCTTGAACCCCGAAGACCTCATACGGCATCGAGTATCGACGAAGCTG 1510  
 Db 391 TGLTLESERPhesapProgluAnleuIlethGlyIleGluValIleaspGluGlyTr 411  
 QY 1511 GTGCGGTGCTATGAGCGCGGATGCGCATTTTGCGATCTTCCCTGCCAATGAGTGAAGCT 1570  
 Db 411 pTPARGLyTgTgTgProaspGlyHisPheGlyMetPheProAlaAntyrValGluLe 431  
 QY 1571 CATTGAG 1577  
 Db 431 uileGlu 433

RESULT 9  
 ID Q8BH56 PRELIMINARY; PRT; 432 AA.  
 AC Q8BH56  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Drosophila-like.  
 GN Drosophila.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N. A.  
 RC STRAIN=C57BL/6J; TISSUE=Brain, Eye, and Medulla oblongata;  
 RX MEDLINE=22334683; PubMed=12466851;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs."  
 RL Nature 420:563-573(2002).  
 RN [2]  
 RP SEQUENCE FROM N. A.  
 RC STRAIN=FVB/N; TISSUE=Colon;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R. L., Feingold E. A., Grouse L. H., Derge J. G.,  
 RA Klausner R. D., Collins F. S., Wagner L., Shenmen C. M., Schuler G. D.,  
 RA Altschul S. F., Zeeberg B., Buetow K. H., Schaefer C. F., Bhat N. K.,  
 RA Hopkins R. F., Jordan H., Moore T., Max S. I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A. A., Rubin G. M., Hong L.,  
 RA Stempleton M., Soares W. B., Bonaldo M. F., Casavant T. L., Scheetz T. E.,  
 RA Brownstein M. J., Uscil T. B., Toshnycki S., Carninci P., Prange C.,  
 RA Rana S. S., Loquellano N. A., Peters G. J., Adamson R. D., Mullany S. J.,  
 RA Bosak S. A., McEwan P. J., McKernan K. J., Malek J. A., Gunaratne P. H.,  
 RA Richards S., Worley K. C., Hale S., Garcia A. M., Gay L. J., Hulyk S. W.,  
 RA Villalón D. K., Wuzny D. M., Sodergren E. J., Lu X., Gibbs R. A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A. C., Shevchenko Y., Bouffard G. G.,  
 RA Blakesley R. W., Touchman J. W., Green E. D., Dickson M. C.,  
 RA Rodriguez M. I., Skalska U., Smalins D. E., Scherch A., Schein J. E.,  
 RA Jones S. J., Mair M. A.,  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [3]  
 RP SEQUENCE FROM N. A.  
 RC STRAIN=FVB/N; TISSUE=Colon;  
 RA Strausberg R.,  
 RA Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.  
 RL EMBL: AK046073; BAC32592.1; -  
 DR EMBL: AK053795; BAC35528.1; -  
 DR EMBL: AK078082; BAC37118.1; -  
 DR EMBL: BC046430; AAH46430.1; -  
 DR MGI: 700006; Dbn1.  
 DR GO: GO:0005737; C:cytoplasm; IDA.  
 DR GO: GO:003027; C:lamellipodium; IDA.  
 DR GO: GO:0003779; F:actin binding; IDA.  
 DR GO: GO:0005515; F:protein binding; IDA.

DR GO: GO:0016601; P:RAC protein signal transduction; IDA.  
 DR InterPro: IPR002108; Actbind\_cofin.  
 DR InterPro: IPR001452; SH3.  
 DR Pfam: PF00241; Cofilin\_ADF; 1.  
 DR Pfam: PF00018; SH3; 1.  
 DR PRINTS: PR00452; SH3DOMAIN.  
 DR PRODOM: PD000066; SH3; 1.  
 DR SMART: SM00102; ADF; 1.  
 DR SMART: SM00326; SH3; 1.  
 DR PROSITE: PS00002; SH3; 1.  
 DR SEQUENCE 432 AA; 48341 MW; 4A36D3B6E59C707F CRC64;  
 SO

Alignment Scores:  
 Pred. No.: 4,58e-43 Length: 432  
 Score: 694.50 Matches: 141  
 Percent Similarity: 69.20% Conservative: 14  
 Best Local Similarity: 62.95% Mismatches: 32  
 Query Match: 16,428 Indels: 37  
 DB: 11 Gaps: 5

US-10-028-952A-4 (1-2333) x Q8BH56 (1-432)

QY 957 AGAGCGCGTGGCCCGCGCGCGGACCGGGGTCGGGTGGAGTGGCCCTTCGCTCG 1016  
 Db 229 ArgSerAlaGlyAlaProArgThrGly----- 237  
 QY 1017 GGAAACGGGCGCGCGGCGGAAAGCGCGCGCCCTCGCCGTCACGACGACGACGCTTCG 1076  
 Db 238 -----GluProGluGluGlu-----AlaValSerArgThrArgGlu 250  
 QY 1077 TGGGGAAC----- 1085  
 Db 251 TPGLuSerAlaGlyGlnGlnAlaProHisProArgGluIlePheGlyGlnLysGluArg 270  
 QY 1086 TGGCGCTAAACCACTCCATCTCCAGTCTCCAGCTCGGACGCTGAGAGCCCTTCCTG 1145  
 Db 271 AlaMetSerThrThrSerValThrSerSerGlnProGlyLysLeuArgSerProPheLeu 290  
 QY 1146 CAGAACGACCTACCCCAACCAAGACCCACTTGGCAGAGACCGACGCTGCTCCATCTCA 1205  
 Db 291 GlnLysGlnLeuThrGlnProGluThrSerThrArgGluArgGluProThrAlaProValSer 310  
 QY 1206 AGGCCAGGCGACATCTCCCTGCTGAGAGCGCGCGCCGACGACCTCCATGCTGGTG 1265  
 Db 311 ArgProAlaAlaGlyVal---CysGluGluProAlaProSerThrLeuSerSer---Ala 328  
 QY 1266 CAGGCGAAGAGAGAGGCTGTATAGAACTCCAGAGCAGAGAACTTTACGAGCAG 1325  
 Db 329 GlnThrGlnGlnGluProThrThrArgValProProGluGlnAspThrLeuThrGluGlu 348  
 QY 1326 CCCCCCTGTCAGAGAGAGAGGCTGCTGAGCATTTGAGCATTTGACACACCATCAGGAC 1385  
 Db 349 ProProLeuValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 368  
 QY 1386 CAGGGCTAGTGGGCGAGGCTGTGTGCGCGCTGCTGACGATTCAGGACGCGGAC 1445  
 Db 369 GlnGlyPheSerGlyGlnGlyLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 388  
 QY 1446 GACACAGATATCTCTTGAACCCGAGAACCTCATCAGGAGGATCGAGGTGATCGAGCA 1505  
 Db 389 AspThrGlnIleSerPheaspProgluAnleuIlethGlyIleGluValIleaspGlu 408  
 QY 1506 GGCTGTGCGCTGATGAGCGCGGATGAGCCATTTTGCGATCTTCCCTGCCAATGAGT 1565  
 Db 409 GlyTTPARGLyTgTgTgProaspGlyHisPheGlyMetPheProAlaAntyrVal 428  
 QY 1566 GAGCTCATGAG 1577  
 Db 429 GluLeuIleGlu 432

RESULT 10  
 ID Q9JHL4 PRELIMINARY; PRT; 436 AA.

AC Q9JHL4; 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE SH3P7R1.  
 GN SH3P7R1.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Yamazaki H., Shirao T.;  
 RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Yamazaki H., Shirao T.;  
 RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 DR EMBL; AB038365; BAA90867.1; -;  
 DR EMBL; AB039346; BAA90819.1; -;  
 DR HSSP; P19174; 2HSP.  
 DR GO; GO:0005622; C:intracellular; IEA.  
 DR GO; GO:0003779; F:actin binding; IEA.  
 DR InterPro; IPRO02108; Actbind\_cotln.  
 DR InterPro; IPRO01452; SH3  
 DR Pfam; PF00241; coiledin\_ADP; 1.  
 DR PRINTS; PR00452; SH3DOMAIN.  
 DR ProDom; PD000066; SH3; 1.  
 DR SMART; SM00102; ADP; 1.  
 DR SMART; SM00326; SH3; 1.  
 DR PROSITE; PS50002; SH3; 1.  
 KW SH3 domain.  
 SQ SEQUENCE 436 AA; 48612 MW; 064CA1E6BAF84444 CRC64;  
 Alignment Scores:  
 Pred. No.: 6,5e-41 Length: 436  
 Score: 665,50 Matches: 138  
 Percent Similarity: 67,11% Conservative: 13  
 Best Local Similarity: 61,33% Mismatches: 35  
 Query Match: 15,73% Indels: 39  
 DB: Gaps: 5  
 US-10-028-952a-4 (1-2333) x Q9JHL4 (1-436)  
 QY 1014 CTGGGAAACGGGGCGCGCGGAAAGCGG-----CCG 1046  
 DB 214 LeuGInGluAlaAlaArgArgGluGlnArgTyGInGluGlnHisArgSerAlaGlyPro 233  
 QY 1047 CCCCCTGCGCCGCTGACGACCC-----GCACGTTGCGGGA 1082  
 DB 234 ProSerProSerSerArgThnGlyGluLeuGluGlnGluValSerArgSerArgGln 253  
 QY 1083 ACCTGGCGC----- 1091  
 DB 254 GluTrpIuSerAlaGlyGlnGlnAlaProHisProArgGluLeuPheLysGlnLysGlu 273  
 QY 1092 -----TAAACCACTCCATCTCCAGCTTCAGCTGGAAGCTGAGAGCCCTTC 1142  
 DB 274 ArgAlaMetSerThrTrpSerValSerSerSerGlnProGlyLysLeuArgSerProPhe 293  
 QY 1143 CTGAGAGAGAGCTCACCAACCAAGAGACCACTTGCGAGAGAGAGCTGTCGACATC 1202  
 DB 294 LeuGlnLysGlnPheTrpGlnProGluAlaSerTyGlnArgGluProThrSerProVal 313  
 QY 1203 TCAAGGCCCAAGGAGATCTCCCTGCTGAGAGAGCCGCGCCAGACATCTCCATGTCG 1262  
 DB 314 SerArgProAlaAlaGlyVal---CyeGluGluLeuAlaProSerThrProSer--- 331  
 QY 1263 GTGAGGCAAGAGAGAGCGCTGTATAGAGAACTCCAGAGAGAGAGAGAGCTTCTAGAG 1322

DB 332 AlGInThrAspAspGluProThrTyGlnValProSerGluGlnIuThrLeuTyGln 351  
 QY 1323 CAGCCCCCACTGAG 1382  
 DB 352 GluProProValGlnGlnProGlyAlaGlySerGlyHisLeuAspThrMetGln 371  
 QY 1383 GCGCAGGAGCTCAGTGGGCAAGGAGCTGAGCCCGCGCCCTGACAGCACTACAGAGCC 1442  
 DB 372 SerGlnAspLeuSerGlnGlnGlyLeuCyAlaArgAlaLeuTyAspTyGlnAlaAla 391  
 QY 1443 GACGACAGAGATCTCTTGAACCCGAGACCTCATCAGGAGATCGAGTGCATGAC 1502  
 DB 392 AspAspThrGluLeuSerPheAspProGluAsnLeuIleThrGlyIleGluValIleAsp 411  
 QY 1503 GAGAGCTGAGCGGTGCTATAGGCGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1562  
 DB 412 GluGlyTrpTrpArgGlyGlyGlyProAspGlyHisPheGlyMetPheProAlaIleTy 431  
 QY 1563 GTGAGGCTCATTGAG 1577  
 DB 432 ValGluLeuIleGlu 436  
 RESULT 11  
 ID Q9JH74 PRELIMINARY; PRT; 429 AA.  
 AC Q9JH74;  
 DT 01-OCT-2000 (Tremblrel. 15, Created)  
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE SH3P7R4 (fragment).  
 GN SH3P7R4.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OC NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Yamazaki H., Shirao T.;  
 RT "Rattus norvegicus mRNA for SH3P7 (SH3P7R4)."  
 RT Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 DR EMBL; AB038364; BAA90866.1; -;  
 DR HSSP; P19174; 2HSP.  
 DR GO; GO:0005622; C:intracellular; IEA.  
 DR GO; GO:0003779; F:actin binding; IEA.  
 DR InterPro; IPRO02108; Actbind\_cotln.  
 DR InterPro; IPRO01452; SH3.  
 DR Pfam; PF00241; coiledin\_ADP; 1.  
 DR PRINTS; PR00452; SH3DOMAIN.  
 DR ProDom; PD000066; SH3; 1.  
 DR SMART; SM00102; ADP; 1.  
 DR SMART; SM00326; SH3; 1.  
 DR PROSITE; PS50002; SH3; 1.  
 KW SH3 domain.  
 FT NON TER  
 SQ SEQUENCE 429 AA; 47923 MW; A7E7E7FBAF4A1705 CRC64;  
 Alignment Scores:  
 Pred. No.: 7,06e-41 Length: 429  
 Score: 665,00 Matches: 136  
 Percent Similarity: 68,64% Conservative: 15  
 Best Local Similarity: 61,82% Mismatches: 35  
 Query Match: 15,72% Indels: 34  
 DB: Gaps: 5  
 US-10-028-952a-4 (1-2333) x Q9JH74 (1-429)  
 QY 1014 CTGGGAAACGGGGCGCGCGGAAAGCGG-----CCG 1049  
 DB 212 LeuGInGluAlaAlaArgArgGluGlnArgTyGlnGluGlnHisArgSerAlaGlyPro 231



DR EMBL; AB039618; BA92708.1; -.  
DR HSSP; P15174; 2HSF.  
DR GO; GO:0005622; C:intracellular; IEA.  
DR GO; GO:0003779; F:actin binding; IEA.  
DR InterPro; IPRO0108; Actbind\_cofin.  
DR InterPro; IPRO01452; SH3.  
DR Pfam; PF00241; Cofilin\_ADF.1.  
DR Pfam; PF00018; SH3.1.  
DR PRINTS; PR00452; SH3DOMAIN.  
DR ProDom; PD000065; SH3.1.  
DR SMART; SM00102; ADF.1.  
DR SMART; SM00326; SH3.1.  
DR PROSITE; PS50002; SH3.1.  
DR SH3 domain.  
KW SEQUENCE 433 AA; 48341 MW; 952FFP81E6AEEDA CRC64;

Alignment Scores:	
Pred. No.:	1,4e+0
Score:	66.00
Percent Similarity:	68.02%
Best Local Similarity:	61.26%
Query Match:	15.63%
BB:	11
Length:	433
Matches:	136
Conservative:	15
Mismatches:	35
Indels:	36
Gaps:	5

US-10-028-952A-4 (1-2333) x Q9JM67 (1-433)

QY	1014	CTGGGAAACGGGCGCGCCGGAAGCGG-----CGC	1046
Db	214	LeuGlnGluAlaAlaArgArgGlnGlnIleArgTyrGlnGlnGlnIleSAhSerAlaGlyPro	233
QY	1047	CCCCCTGCGCCG-----TCACGACCCGACCGCTTCGTGGGA	1082
Db	234	ProSerhArgThrGlyGlnLeuGlnGlnGlnValValSerAlaArgSerArgGlnGlnTyrGlu	253
QY	1083	ACC-----TGCGCG	1091
	:::		
Db	254	SerAlaGlyGlnGlnAlaProHlaProArgGlnIlePheLysGlnLysGluAlaMet	273
QY	1092	TAAACACCTCCATCTCCAGTCCACGCTGTGGCAAGCGAGAGCCCTTCTCTGACAAAG	1151
Db	274	SerThrThrSerValSerSerSerGlnProGlyLysLeuAlaGserProPheLeuGlnLys	293
QY	1152	CAGCTCACCCACCAAGAGACCACTTTGGCAGAGAGCCACTGCTGCATCTCAAGGCC	1211
Db	294	GlnPheThrGlnProGlnAlaSerTyrGlyArgGlnProThrSerProAlaSerArgPro	313
QY	1212	AGGGCAGATCCTCCGCTGAGAGAGCGCGCCACAGACTCTCCATGCTGTGGTCAGAGCA	1271
Db	314	AlaAlaGlyVal---CysGlnGlnLeuAlaProSerThrProProSer---AlaGlnThr	331
QY	1272	GAAGAGAGGCGTGTGTATGAGAACTTCCAGACGAGAGACTTTCAGAGACGCCCA	1331
Db	332	AspAspGlnProThrTyrGluValProSerGlnGlnGlnTyrLeuTyrGlnGluProPro	351
QY	1332	CTGTGTGACAGCAAGCAAGGTCGTGGCTGTAGCAATTGACACCAATTCAGAGGCGAGGG	1391
Db	352	ProValGlnGlnProGlyAlaGlySerGlyHisIleAspAspTyrMetGlnSerGlnAsp	371
QY	1392	CTCAGTGGGCAAGGCGCTGTGTCCCGTGCCTGTACGAATACAGACGAGCCGACGACCA	1451
Db	372	LeuSerGlyGlnGlyLeuCysAlaArgAlaLeuTyrAspTyrGlnAlaAlaAspAspThr	391
QY	1452	GAGATCTCTTTGACCCCGAAGAACTCATACAGGGCATCGAGGTATGAGAGGAGCGTG	1511
Db	392	GluIleSerPheAspProGlnAsnLeuIleThrGlyIleGluValIleAspGluIleTyr	411
QY	1512	TGGCGTGGCTATGGGCGCGGATGACATTTGGCATTTCCCTCCCACTACGTGAGAGCTC	1571
Db	412	TyrAlaGlyTyrGlyProAspGlyHisPheGlyMetPheProAlaAsnTyrValGlnLeu	431
QY	1572	ATTGAG	1577
Db	432	IleGlu	433

RESULT 14

Q7ZXQ9	PRELIMINARY;	PRT;	447 AA..
ID Q7ZXQ9			

```

AC Q72X09
DT 01-JUN-2003 (T-EMBLrel. 24, Created)
DT 01-JUN-2003 (T-EMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)
DE Similar to drebrin-like.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipiloidea; Pipidae;
OC Xenopodinae; Xenopus.
CX NCBI_TaxID=8355;
RX [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Embryo;
RA Klein S., Strausberg R.,
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC044296; AAH44296.1; -.
DR GO; GO:0005622: Cytoskeletal; IEA.
DR GO; GO:0003773; F-actin binding; IEA.
DR InterPro; IPR002108; Actbind_coflin.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00241; Cofilin_ADF; 1.
DR Pfam; PF00018; SH3_1.
DR PRINTS; PR00452; SH3DOMIN.
DR Prodom; PD000066; SH3_1.
DR SMART; SM00102; ADF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PSS0002; SH3_1.
SQ SEQUENCE 447 AA; 50611 MW; 3BC6767EE6E3B521 CRC64;

```

Alignment Scores:	
Pred. No.:	1.17e-20
Score:	392.50
Percent Similarity:	59.13%
Best Local Similarity:	48.55%
Query Match:	9.48%
DB:	13
Gaps:	
length:	447
Matches:	66
Conservative:	15
Mismatches:	48
Indels:	23
Gaps:	6

US-10-028-952A-4 (1-2333) x Q7ZXQ9 (1-447)

```

QY      1110 AGTCCTCAGCTGGCGAAGTGAAGAGCCCTCTCTGCGAAGACAGTCACTCCCAACAAG   1169
        ||| ||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db      273 SerIaGlnProGIlyArgLeuArgSerProheNeuGlnIySeraIaCySgInProGlu   292
        ||| ||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY      1170 ACC-----CACTTGCGAGAAGCCAAGTGTGCTGCATCTCAAG   1208
        :: :||| :||| :||| :||| :||| :||| :||| :|||
Db      293 SerSerProProPserProValHisArgValGlnGluProPserProProValTy   312
        ||| ||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY      1209 CCCAGGCGAGATCTCCCTGCTGAGAGCGCGGCC-----AGCATCTCTCA-----   1256
        ||| ||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db      313 ProAlaHisGlnThrProProGluSerProValProProValSerHisProProGluSer   332
        ||| ||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY      1257 -----TGCTGTGTG---CAGCGAGAAGAGAGCTGTGTAATG   1297
        ||| ||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db      333 ThrValHisValLySgGluGlnCyenThrLaSerGlnGlnGluGlnIyAsnIleTyGln   352
        ||| ||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY      1299 GAACCTCCAGACGAGGAAACCTTTACAGAGAGCCCCACAGTGGTCAGACAG-----CAA   1346
        ::| :||| :||| :||| :||| :||| :||| :|||
Db      353 AspaIaThrGluAspGlnAsnIleTyGlnIyspThrThrGluAsnGlnAsnIleTyGln   372
        ||| ||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY      1347 GGTGTGCTGCTGAGACACATTGACCACACATTCAGAGGCGAGGGCTCAGTGGCGCAGG   1406
        ||| ||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
Db      373 AspThrProGlnGluGluProValTyGlnIleGluValGlu-----GluGluIySgLy   390
        ||| ||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY      1407 CTCGTGCCCCTGCTCCCTGATGACTTACACGAGCACCCGACACACAGATACTCCCTTGAC   1466
        ::| :||| :||| :||| :||| :||| :||| :|||
Db      391 ValIcySalaArgIaleuTyIyrAspTyGlnAlaIaaPaapThrGlnIleSerPheasp   410
        ||| ||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||
QY      1467 CCGGAAACCTCATACGCGGCAATCAGAGTGATCGACGAAGCTGTGTGCTGTGCTATGAG   1526
        ||| ||||| :||||| :||||| :||||| :||||| :||||| :||||| :|||||

```

Db 411 Probsapleuilethrginlelelnpneileaspeluglytrprrparglypser 430

QY 1527 CCGATGCGCATTTTGCGATGTTCCCTGCCACCTAGCTGAGCTCATTTAG 1577

Db 431 ProalaglylispheglymetpneProalaaantyValglueneuglu 447

RESULT 15

Q9XUT0 PRELIMINARY; PRT; 643 AA.

AC Q9XUT0; 01-NOV-1999 (TREMBLrel. 12, Created)

DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)

DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)

DE K08E3.4 protein.

GN K08E3.4.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoea; Chromadorea; Rhabditida; Rhabditioidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

OX [1]

RP SEQUENCE FROM N.A.

RA McMurray A.A.;

RL Submitted (NCV-1996) to the EMBL/GenBank/DBJ databases.

RN [2]

RP MEDLINE=99069613; PubMed=9851916;

RT "Genome sequence of the nematode C.elegans: A platform for

RT investigating biology."

RT Science 282:2012-2018(1998).

CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.

DR EMBL; Z81568; CAB04592.1; -

DR PIR; T23453; T23453.

DR HSSP; P29355; 1SEM.

DR WormPep; K08E3.4; CE18868.

DR GO; GO:0005622; C:intracellular; IEA.

DR GO; GO:0003779; F:actin binding; IEA.

DR InterPro; IPR002108; Actbind\_cofin.

DR InterPro; IPR001452; SH3.

DR Pfam; PF00241; Cofilin\_ADF; 1.

DR PRINTS; PR00018; SH3; 1.

DR PRINTS; PR00452; SH3DOMAIN.

DR PRODOM; PD000066; SH3; 1.

DR SMART; SM00326; SH3; 1.

DR PROSITE; PS50002; SH3; 1.

DR SH3 domain.

QW SEQUENCE 643 AA; 71941 MW; B95BF648425EP694 CRC64;

Alignment Scores:

Pred. No.: 6,35e-13 Length: 643

Score: 288.50 Matches: 92

Percent Similarity: 38.04% Conservative: 32

Best Local Similarity: 28.22% Mismatches: 105

Query Match: 6.82% Indels: 97

DB: 5 Gaps: 11

US-10-028-952a-4 (1-2333) x Q9XUT0 (1-643)

QY 841 GCCCAGGCGAAGCATACGCGAGCGCGGAGCTCGTGGCTTCGATGACCGGATCC 900

Db 335 AlaglnlysglnygluphegluthrasparglyshrgllelaaspsapleuHlsgly 354

QY 901 CCGGCTGTCCCGCGCGGCGGCGCGCCCGCCCGACGCGCGCGCGGAGGAG 960

Db 355 lyshlaaglnlleeseraplyspovalprosersevalasnproala----- 370

QY 961 CCGGTGCGCGCGCGCGCGGAGCGCGGAGTCCGATGCGGAGTCCCTTGCTTGGAA 1020

Db 371 -----ProalaproalaprothrProserAlaglyleu---Valglyserarglysglu 387

QY 1021 ACGGGGGCGCGCGGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1080

Db 388 leupheseriserysProserglyProvalleuProlys-----Proglinalaengly 405

QY 1081 GAACCTGGCGGCTAAACCACTTCATCTTCAGTCTCAGCTTGGGAGAGAGCCCT 1140

Db 406 serProlylsystrpProPro--ValglythrtrserProPro-----ArggluPro 423

QY 1141 TCCTGCAGAGAGAGGTCCACCCACGAG----- 1179

Db 423 al-----AsnargylethrgluProasgluproserThyThrProlysProilea 441

QY 1180 GCGAGAGCCGAGCTGCTGCCATCTCAAGGCCGAGGAGATCTCCCTGCT----- 1229

Db 441 lATygluProgluprometValTyrysProglualametylsProalavalserTyra 461

QY 1230 -----GAGGAGCGGCGCCGAGCACT----- 1250

Db 461 spAlaTyrgluGlupProProalagluProalaprothrPheleuAlaprothrPro 481

QY 1251 -----CCTCCATGCTGTGTCAGGAGAAAGAGAGGCTGTATG 1290

Db 481 alIleAlaProProProProglupProthrProAlaProserHisTyralaserGlnTyra 501

QY 1291 AGGAACCTCCAGAGAGAGACCTTC----- 1316

Db 501 spAlaProProvalHisgluIseryPhegluProalavalProProvalserAlaproserH 521

QY 1317 -----TACGAGCAGCCCGCACTGTGTCAGCAGCAGAGAGTGGCTGTGAGC 1362

Db 521 lsetyAlaserGlnTyrsAlaproProglupProIleaspseryHisererserSerg 541

QY 1363 ACATTGACCAACCACTTAGGCGCAG----- 1388

Db 541 InleuProAlaHisIleAlaIseryGlnTyraspmetProProvalmeProgluInuPro 561

QY 1388 ----- 1388

Db 561 alpheaAlaProlyserserProIlelysValAlaAlaProProIleaspInTyrasp 581

QY 1389 -----GGGCTCAGTGGGAGAGGCTCTGTGCCCGGCTTACGACTTACGAGCAG 1440

Db 581 heProProAlaValAlaIleGlnIns-----AlametyAlaLeuTrpAspyGlnAla 599

QY 1441 CCGAGCAGCAGAGATCTCTTGAACCGCGAAGCTCATCAGGAGCATCGAGGTATCG 1500

Db 599 lAspsapThrGlnIleeserPheaspproAlasppilleIleThrAspIleaspeInVala 619

QY 1501 ACGAAGCTGTGCGGTATGAGCGCGATGCGCATTTTGGATGTCTCCCTGCCAACT 1560

Db 619 spSerGlyTrpTrpIlyeslylYargAlaProserGlylYargValglyLeuPheProAlaant 639

QY 1561 ACGTGAGCTCATT 1574

Db 639 yTVallylsleuile 643

Search completed: April 1, 2004, 15:58:20

Job time : 139 secs



Fri Apr 2 07:35:04 2004

us-10-028-952a-4.rsp

Page 1

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 1, 2004, 15:47:06 ; Search time 25 Seconds  
(without alignments)  
9718.374 Million cell updates/sec

Title: US-10-028-952A-4  
Sequence: 1 gcacgagatccaccatcgtcc.....agcttgagacttcgcga 2333

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 283362

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+n2p.model -DEV=xip  
-O=/cgn2.1/USPTO.spool.p/US10028952/runat.01042004.154625.643/bpp.query.fasta.1.2503  
-DB=SwissProt 42 -QFWT=fastan -SUFFIX=rsf -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdt -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFW=pic -NORW=ext -HEAPSIZ=500 -MINLEN=0 -MAXLEN=200000000 -ICPU=3  
-USFR=US10028952.@CGC.1.1.25@runat.01042004.154625.643 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DEPBLOCK=100 -KONJUG  
-DEV\_TIMEOUT=120 -MARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FCGAPOP=6  
-FCGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : SwissProt\_42.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	277	6.5	486	HS1_HUMAN	P14317 homo sapien
2	261.5	6.2	486	HS1_MOUSE	P49710 mus musculu
3	244.5	5.8	563	SRGB_CHICK	Q01406 gallus gall
4	242.5	5.7	660	YHL1_EBV	P01181 Epstein-bar
5	231	5.5	546	SRGB_MOUSE	Q60598 mus musculu
6	230	5.4	550	SRGB_HUMAN	Q14247 homo sapien
7	214	5.1	857	APPI_SCHPO	O9P768 schizosacch
8	195.5	4.6	5262	MLL2_HUMAN	O14656 homo sapien
9	193.5	4.6	825	SE5_RAT	O63003 rattus norv
10	191.5	4.5	660	YHL1_EBV	P01311 Epstein-bar
11	191	4.5	775	ICPO_HSV11	P08393 herpes simp
12	189	4.5	1455	CA11_CHICK	P02457 gallus gall
13	186.5	4.4	1355	CA21_RANCA	Q42350 ratius norv
14	185.5	4.4	1183	DRPL_RAT	O54258 rattus norv
15	184.5	4.4	1336	Y146_HUMAN	O9C018 homo sapien
16	184.5	4.4	1453	CA11_MOUSE	P11087 mus musculu
17	182	4.3	592	ABP1_YEAST	P15891 saccharomyc
18	181.5	4.3	1355	CA21_RANCA	O42350 ratius norv

19	180.5	4.3	1426	1	BCL9_HUMAN	O00512 homo sapien
20	180.5	4.3	1460	1	CA11_CANFA	O9X617 canis famli
21	180	4.3	617	1	ABP1_SACEX	P38479 saccharomyc
22	180	4.3	1262	1	CA13_CHICK	P12105 gallus gall
23	178.5	4.2	1372	1	CA13_MOUSE	Q01149 mus musculu
24	178	4.2	699	1	VGLG_HSV2H	P13290 herpes simp
25	177.5	4.2	1464	1	CA13_MOUSE	P08121 mus musculu
26	175.5	4.2	1466	1	CA13_HUMAN	P02461 homo sapien
27	175.5	4.1	2716	1	OSA_DROME	O81n94 drosophila
28	175	4.1	3530	1	MY15_HUMAN	O9ukn7 homo sapien
29	174.5	4.1	1003	1	MYD6_HUMAN	O96dn6 homo sapien
30	174	4.1	1185	1	DRPL_HUMAN	P54255 homo sapien
31	173.5	4.1	1461	1	IE18_PRIVIF	P11675 pseudorabie
32	172.5	4.1	1464	1	CA11_HUMAN	P02451 homo sapien
33	172	4.1	439	1	XP2_XENLA	P17437 xenopus lae
34	172	4.1	1459	1	CA12_MOUSE	P28481 mus musculu
35	172	4.1	1321	1	TRR2_MOUSE	P31485 pseudorabie
36	171.5	4.1	1321	1	SRK1_HUMAN	P03566 homo sapien
37	171.5	4.1	2161	1	MLL4_HUMAN	Q9umt6 homo sapien
38	171.5	4.1	825	1	ICP0_HSV2H	P28284 herpes simp
39	171	4.0	825	1	ICP0_HSV2H	P08123 homo sapien
40	170.5	4.0	1366	1	CA21_BOVIN	P02453 bos taurus
41	170	4.0	825	1	ICP0_HSV2H	P28284 herpes simp
42	170	4.0	825	1	ICP0_HSV2H	O46393 canis famli
43	170	4.0	2716	1	CA21_CANFA	O81n94 drosophila
44	170	4.0	2716	1	MD6_HUMAN	O96dn6 homo sapien
45	168.5	4.0	1003	1	MD6_HUMAN	O96dn6 homo sapien

# ALIGNMENTS

RESULT 1	
HS1_HUMAN	
ID HS1_HUMAN	STANDARD; PRT; 486 AA.
AC P14317;	
DT 01-JAN-1990 (Rel. 13, Created)	
DT 01-JAN-1990 (Rel. 13, Last sequence update)	
DT 15-MAR-2004 (Rel. 43, Last annotation update)	
DE Hematopoietic lineage cell specific protein (hematopoietic cell-specific LYN substrate 1) (LCKBP1).	
DE HCLSI OR HS1.	
OS Homo sapiens (Human).	
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
OC NCBI_Taxid=9606;	
RN [1]	
RP SEQUENCE FROM N.A.	
RX MEDLINE=90067934; PubMed=2587259;	
RA Kitamura D., Kaneko H., Miyagoe Y., Ariyasu T., Watanabe T., Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marusita K., Farmer A.A., Rubin G.M., Hong L., Stalstein M., Soares W.B., Bonaldo M.F., Casavant T.L., Schetz T.E., Brownstein W.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Lequellain N.A., Peters G.J., Abramson R.D., Mullany S.J., Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gamaratte P.H., Villalón D.K., Muzny K.C., Hale S., Garcia A.M., Gay L.J., Halys S.W., Richards S., Morley K.C., Hale S., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Heiton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman U.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W., Butlerfield Y.S.N., Krzyzinski M.I., Skalska U., Smallus D.E., Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;	

RT Generation and initial analysis of more than 15,000 full-length  
RT human and mouse cDNA sequences.";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16999-16903 (2002).  
RN [3]  
RP SEQUENCE OF 97-108; 193-201 AND 240-248.  
RX MEDLINE=96311348; PubMed=8713105;  
RA Egerston M., Moritz R.L., Draker B., Kelso A., Simpson R.J.,  
RT "identification of the 70kd heat shock cognate protein (Hsc70) and  
RT alpha-actinin-1 as novel phosphotyrosine-containing proteins in T  
RT lymphocytes.";  
RL Biochem. Biophys. Res. Commun. 224:666-674 (1996).  
RN [4]  
RN BINDING TO HAX-1 PROTEIN.  
RX MEDLINE=9721181; PubMed=9058808;  
RA Suzuki Y., Demoliere C., Kitamura D., Takeshita H., Deuschle U.,  
RN Matanabe T.,  
RT "HAX-1, a novel intracellular protein, localized on mitochondria,  
RT directly associates with Hsl, a substrate of Src family tyrosine  
RT kinases.";  
RL J. Immunol. 158:2736-2744 (1997).  
RN [5]  
RN PHOSPHORYLATION OF TYR-222.  
RX MEDLINE=9916757; PubMed=10066823;  
RA Brunati A.M., Donella-Deana A., James P., Quadroni M., Contri A.,  
RA Marin O., Pina L.A.,  
RT "molecular features underlying the sequential phosphorylation of Hsl  
RT protein and its association with c-Fgr protein-tyrosine kinase".  
RL J. Biol. Chem. 274:7557-7564 (1999).  
CC -1- FUNCTION: Substrate of the antigen receptor-coupled tyrosine  
CC kinase. Plays a role in antigen receptor signaling for both  
CC clonal expansion and deletion in lymphoid cells. Directly  
CC associates with HAX-1, through binding to its C-terminal region.  
CC May also be involved in the regulation of gene expression.  
CC -1- SUBUNIT: Associates with the SH2 and SH3 domains of LCK.  
CC -1- SUBCELLULAR LOCATION: Mitochondrial (Probable).  
CC -1- TISSUE SPECIFICITY: Expressed only in tissues and cells of  
CC hematopoietic origin.  
CC -1- DEVELOPMENTAL STAGE: Expressed in early stage of myeloid and  
CC erythroid differentiation.  
CC -1- PTM: Phosphorylated by LYN; rapidly after crosslinking of surface  
CC IgM on B cells.  
CC -1- SIMILARITY: TO CHICKEN P80/85 PROTEINS (CORTACTIN).  
CC -1- SIMILARITY: Contains 1 SH3 domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL institution.  
CC The European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: X16663; AA34651.1; -.  
DR EMBL: BC016758; AA16758.1; -.  
DR PIR: S07633; S07633.  
DR HSSP: P29355; ISEM.  
DR GENEW: HGNC:4844; HCLSL1.  
DR MIM: 601306; -.  
DR GO: GO:0005655; C:DNA-directed RNA polymerase II, core complex, TAS.  
DR GO: GO:0003700; P:transcription factor activity; TAS.  
DR GO: GO:0007242; P:intracellular signaling cascade; TAS.  
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; TAS.  
DR InterPro: IPR001133; Hsl/Cortactin.  
DR InterPro: IPR00108; Neu\_cyc\_fac2\_2.  
DR InterPro: IPR001452; SH3.  
DR Pfam: PF02218; HSL\_rep; 4.  
DR Pfam: PF00018; SH3; 1.  
DR PRINTS: PRC0499; P67PHOX.  
DR PRINTS: PRC0452; SH3DOMAIN.  
DR PRODOM: PD000066; SH3; 1.  
DR SMART: SMO0326; SH3; 1.  
DR PROSITE: PS50002; SH3; 1.  
DR Repeat: SH3 domain; Phosphorylation.  
KW

[illegible]

ID	HS1_MOUSE	STANDARD	PRT:	486 AA.
AC	P49710			
DT	01-OCT-1996	(Rel. 34, Created)		
DT	01-OCT-1996	(Rel. 34, Last sequence update)		
DT	15-MAR-2004	(Rel. 43, Last annotation update)		
DE	Hematopoietic lineage cell specific protein (Hematopoietic cell-specific LYN substrate 1) (LCBP1).			
DE	HCUS1 OR HSL.			
CN	Mus musculus (Mouse).			
OS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrate; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_Taxid=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
FX	MEDLINE=95217198; PubMed=7535527;			
FA	Kitamura D., Kaneko H., Taniuchi T., Yamamura K.I., Watanabe T.;			
RT	"Molecular cloning and characterization of mouse HSL.";			
RL	Biochem. Biophys. Res. Commun. 208:1137-1146(1995).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
FX	STRAIN=BA1B/C;			
FX	MEDLINE=97098949; PubMed=8943564;			
RA	Takemoto Y., Sato M., Furuta M., Hashimoto Y.;			
RT	"Distinct binding patterns of HSL to the Src SH2 and SH3 domains			
RT	reflect possible mechanisms of recruitment and activation of			
RT	downstream molecules.";			
RL	Int. Immunol. 8:1699-1705(1996).			
CC	-1- FUNCTION: Substrate of the antigen receptor-coupled tyrosine			
CC	kinase. Plays a role in antigen receptor signaling for both clonal			
CC	expansion and deletion in lymphoid cells. Directly associates with			
CC	HAX-1, through binding to its C-terminal region. May also be			
CC	involved in the regulation of gene expression (By similarity).			
CC	-1- SUBUNIT: Associates with the SH2 and SH3 domains of LCK.			
CC	-1- SUBCELLULAR LOCATION: Mitochondrial (By similarity).			
CC	-1- TISSUE SPECIFICITY: Expressed only in tissues and cells of			
CC	hematopoietic origin.			
CC	-1- PTM: Phosphorylated by LYN; rapidly after crosslinking of surface			
CC	IGM on B cells (By similarity).			
CC	-1- SIMILARITY: TO CHICKEN P80/85 PROTEINS (CORTACTIN).			
CC	-1- SIMILARITY: Contains 1 SH3 domain.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on ways			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).			
CC	-----			
DR	EMBL; D42120; BAA07701.1; -			
DR	EMBL; X84797; CAA59265.1; -			
DR	PIR; I49760; I49760.			
DR	HSSP; P29355; ISEM.			
DR	MGD; MG1:104568; HCLs1.			
DR	GO; GO:0017124; F:SH3-domain binding; IDA.			
DR	InterPro; IPR003134; Hsl/Cortactin.			
DR	InterPro; IPR001452; SH3.			
DR	InterPro; IPR001452; SH3.			
DR	Pfam; PF02218; Hsl_rep; 4.			
DR	Pfam; PF00018; SH3; 1.			
DR	PRINTS; PR00489; P67PHOX.			
DR	PRINTS; PR00452; SH3DOMAIN.			
DR	PRODom; PD000066; SH3; 1.			
DR	SMART; SM00326; SH3; 1.			
DR	PROSITE; PS50002; SH3; 1.			
KW	Repeat; SH3 domain; phosphorylation.			
FT	DOMAIN 27 66			
FT	INVOLVED IN HAX-1 BINDING (BY			
FT	SIMILARITY).			
FT	DOMAIN 81 214			
FT	REPEAT 81 116			
FT	REPEAT 117 153			
FT	2.			

[illegible]

Db 470 Gln---CysargglyHisphneglyLeupheProAlaSerntlyVallyshleuou 486

RESULT 3

SRCS\_CHICK STANDARD; PRT; 563 AA.

ID SRC8\_CHICK

AC 001406;

DT 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Src substrate protein p85 (p80) (Cortactin).

GN EWS1 OR P85.25

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

OC NCBI\_TaxID=9031;

OX [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92017786; PubMed=1922035;

RA W.H., Reynolds A.B., Kanner S.B., Vines R.R., Parsons J.T.;

RT Identification and characterization of a novel

RT cytoskeleton-associated p60src substrate."

RL Mol. Cell. Biol. 11:5113-5124(1991).

CC -1- FUNCTION: May contribute to the organization of cell structure. The SH3 motif may function as a binding region to cytoskeleton. Tyrosine phosphorylation in transformed cells may contribute to cellular growth regulation and transformation.

CC -1- SUBCELLULAR LOCATION: In normal cells, probably in association with the plasma membrane and possibly the endoplasmic reticulum. p80/85 colocalizes with F-actin in peripheral extensions of normal cells and rosettes (podosomes) of src-transformed cells. p80/85 probably associates with components of the cytoskeleton.

CC -1- ALTERNATIVE PRODUCTS:

CC Event=Alternative initiation;

CC Comment=2 isoforms, p85 (shown here) and p80, may be produced by alternative initiation;

CC -1- PTM: In normal cells, appears to be phosphorylated on serine and threonine; in cells expressing activated forms of pp60-src, they become heavily phosphorylated on tyrosine in vitro.

CC -1- SIMILARITY: TO HEMATOPOIETIC LINEAGE CELL SPECIFIC PROTEIN.

CC -1- SIMILARITY: Contains 1 SH3 domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; M73705; AAA49031.1; -

DR PIR; A41530; A41530.

DR HSSP; P19174; 2HSP.

DR InterPro: IPR003134; Hsl/Cortactin.

DR InterPro: IPR000108; Neu\_cyt\_fact\_2.

DR InterPro: IPR001452; SH3.

DR Pfam; PF02218; HSL\_rep7.

DR Pfam; PF00018; SH3; 1.

DR PRINTS; PR00459; P67PHOX.

DR PRINTS; PR00452; SH3DOMAIN.

DR ProDom; PD000066; SH3; 1.

DR SMART; SMO0326; SH3; 1.

DR PROSITE; PS50002; SH3; 1.

DR Phosphorylation; Repeat; SH3 domain; Cytoskeleton;

KM Alternative initiation.

KM CHAIN 1

FT CHAIN 1 563 SRC SUBSTRATE PROTEIN P85, ISOFORM P85.

FT INIT MET 11 SRC SUBSTRATE PROTEIN P85, ISOFORM P80.

FT DOMAIN 92 338 6.5 X 37 AA TANDEM REPEATS.

FT REPEAT 92 128 1.

FT REPEAT 129 165 3.

FT REPEAT 166 202 3.

FT REPEAT 203 239 4.

FT REPEAT 240 276 5.

FT REPEAT 277 313 6.

FT REPEAT 314 338 7 (INCOMPLETE).

FT DOMAIN 505 563 SH3.

SO SEQUENCE 563 AA; 63329 MW; 22A824A08BD654C CRC64;

Alignment Scores:

Pred. No. 1.05e-06 Length: 563

Score: 244.50 Matches: 66

Percent Similarity: 54.82% Conservative: 25

Best Local Similarity: 39.76% Mismatches: 54

Query Match: 5.78% Indels: 21

DB: 1 Gaps: 6

US-10-028-952a-4 (1-2333) x SRC8\_CHICK (1-563)

QY 1096 CCACCTCCATCTCCAGCTGCTCAGCTGCAAGCTGAGAGCCCTTCTGCAAGAGC 1155

DB 411 ProProserPro-ThrThrgin-----ProAlaGluProly 423

QY 1156 TCACCCCAACCAAGACCCACTTGGCAGAGAGCCAGCTGCTGCATCTCAAGGCCAGG 1215

DB 423 sThrProser---SerProvallyrGlnAspAlaValserTyrAspAlaGluSerAlaTy 442

QY 1216 CAGATCTCCCTGCTGAGAGAGCCGCCACCACTCTCCATGCTGTCAGGACAGAG 1275

DB 442 rlyAsnserSerThrThryserAlaGluHlaGluProGluSerGlyTyrlyThrTh 462

QY 1276 AGAGAGCTGTATAGAGAACCTTCAAGCAG---GAGACTTCTAGAGAGCCCCAC 1332

DB 462 rGly-SerAspTyrGlnAlaValserGlnArgGlnAlaGluTyrGluProGluThrv 482

QY 1333 TGGTCAGACAGAGAGCTGCTGCTGAGACATTCACCACTTCAAGGAGGAGG 1392

DB 482 alyrGlnAlaGlyAlaGly-----AspHisTyrGlnAlaGluGluLent 498

QY 1393 TCAGTGGCAA-----GGCTCTGTGCGCTGCTGCTGAGACTACAGG 1437

DB 498 hTyrAspGluTyrGlnAsnGluLeuAllylethraAlaIleAlaLeuTyrAspTyrGlnA 518

QY 1438 CAGCCGACAGACAGAGATCTCTTGACCCCGAGAACTCATGACGGGATCGAGGTGA 1497

DB 518 laAlaGlyAspAspGluIleSerPheAspProAspAlaIlelethraAsnIleGluActi 538

QY 1498 TCAGCAAGGCTGTGCTGCTGCTGATGAGCCGATGCGCATTTTGGCATGTCCTGCCA 1557

DB 538 leAspAspGlyTyrTrpAlaGlyVal---CysTyrGlyArgTyrGlyLeuPheProAlaA 557

QY 1558 ACTACGTGAGAGCTC 1571

DB 557 snlyrValGluLeu 561

RESULT 4

YH1L\_EBV STANDARD; PRT; 660 AA.

ID YH1L\_EBV

AC P03181;

DT 21-TUL-1986 (Rel. 01, Last sequence update)

DT 21-TUL-1986 (Rel. 01, Last sequence update)

DT 15-TUL-1998 (Rel. 36, Last annotation update)

DE Hypothetical EBV1 protein.

OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).

OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae; Gammaherpesvirinae; Lymphocryptovirus.

OC NCBI\_TaxID=10377;

OX [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=84270667; PubMed=6087149;

RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J., Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C., Tuffnell P.S., Watnall B.G.;

RA "DNA sequence and expression of the B95-8 Epstein-Barr virus genome."

RL Nature 310:207-211(1984).

CC -----  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ch/announce/>  
 CC or send an email to [license@ebi-sib.ch](mailto:license@ebi-sib.ch)).  
 CC -----

DR EMBL; V01555; -; NOT\_ANNOTATED\_CDS.

PIR; A03742; Q08B3.

KM Hypothetical protein; Early protein; Repeat.

FT DOMAIN 149 648 4 X 125 AA TANDEN REPEATS.

FT REPEAT 149 273 1.

FT REPEAT 274 398 2.

FT REPEAT 399 523 3.

FT REPEAT 524 648 4.

SO SEQUENCE 660 AA; 66244 MW; 86DA1D67A37152A2 CRC64;

# Alignment Scores:

Pred. No.:	1.36e-06	Length:	660
Score:	242.50	Matches:	201
Percent Similarity:	30.67%	Conservative:	41
Best Local Similarity:	25.48%	Mismatches:	237
Query Match:	5.74%	Indels:	310
DB:	1	Gaps:	43

US-10-028-952a-4 (1-2333) x YHL\_EBV (1-660)

QY 2186 AGGGTGGAGCTTGGCCATCTGCATCCCACTTCCCACTAGCC--CAGGCAAGGAGACAG 2130  
 Db 47 GlnValAlaLeuAlaAspAlaIleSerProAspLeuGlnIleAspGlnIleSerGlyAla 66  
 QY 2129 AGAATGAGTGGAGGAGCAGACAGAGCTCCAAACAAGCAAAATTCCTGCTGCCAAAC 2070  
 Db 67 ArgProGlyGlyGlyAsn----- 72  
 QY 2069 ACCATGATCCACTGACTTGGTCACAAACTGTGTAACAAATTCCTAGCTTCACT 2010  
 Db 73 -----Arg 73  
 QY 2009 GTTCCCAAGGCTATCTTAACAGATGGGACACACCCCTTAGAAGCAAGAAAGA 1950  
 Db 74 ValAlaIleAlaGlyArgGlyProGlyThrProAlaProSerArgIleSerArgGlyThr 93  
 QY 1949 AAAAAGCAATTAATGACAGAGGACAGCAGAGCTCCAGATGCCCTGCTCAGACCC 1890  
 Db 94 GlyProAlaGlnGln-AlaSerHisIleAlaIleSerAsnProThrGlyGlyCysSerAspPr 113  
 QY 1889 T-----TCCCGGAG-----CTTGGCAGTGGCCATCAGGGCT 1857  
 Db 113 OGlnArgSerProArgThrArgGlnAlaGlyTyrAlaLeuGlyGlyGlySerAlaGlyLe 133  
 QY 1856 TGCGCACAGCAGGAGCTTGGCTCAGTATCCTGTGACAGGAGGAGGAGCCAGCTGTCTGGGA 1797  
 Db 133 uGlySerArg--GlyProArgProHisProAla-----PheIleValGlnTrpse 149  
 QY 1796 GGGTGGGGGATGACAGAGAGATGTGGGAAATACACAGGCCATGTGCTGATTTGGGGTG 1737  
 Db 149 TAlaArg-----AsnProGlyCysProArgThrTrpArg-----ArgArgSerGlyAla 165  
 QY 1736 ACAGAGCTGAGTCTGACAGAGGAGGAGCCCTGAGGCTCAT--CCTCACTGGGGGT- 1681  
 Db 165 agIleArgGlyHisProProProGlyAlaGlyGlnArgProSerGlyPro-ThrGlyGlyAla 185  
 QY 1680 -----CCTATTCTCTGGAAGAGTCTGA 1659  
 Db 185 TGPProAlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGlyGlyAlaAla 205  
 QY 1658 ATGTCACTCCAGAGCTCTCTCCAGACATAAGAGACCATGTCTGAGAGGGAGGG 1599  
 Db 205 IalValProSerGlyAlaIleThrProHisProGlnArgGlySerGlyProAlaAspProPro 225

QY 1598 CAAGATGTCCTTCAAGCTCTCAATGAGCTCCACGTAAGTTGGAGAGA----- 1549  
 Db 225 IalAlaIleArgLeuProProGlnArgGlnGlnIleProAlaLeuProIleAspLeuAlaAla 245  
 QY 1548 -----ACATGGCCAAATATGCG--CATCCGAGCCCATAGCCACAGCCGCTT----- 1504  
 Db 245 IagIleArgCysProAlaGlyProProProThrArgSerGlyAlaAlaIleAlaGlnArgThrH 265  
 QY 1503 --CGTCGATCACTTGAGTCCCGCTGATAGAGTTCTCGGGGGTCAAGAGATCTCTGT- 1447  
 Db 265 IAsArgArgProProGlyCysPro-----ArgSerAlaArgAsnProGlyCysP 281  
 QY 1446 -----CGTGGCTGCTGGTGTAGTGTGACAGGAGCAGC----- 1417  
 Db 281 roArgThrThrArgArgArgSerGlyAlaGlnArgGlyHisProProProGlyAlaGlyG 301  
 QY 1416 -----GGCACAGAGCCCTTGGCCACTGAGCCCTTGGC----- 1384  
 Db 301 IlnArgProSerGlyProThrArgGlyArgProAlaAlaProGlyAlaProGlyThrPro 321  
 QY 1383 -----CTGAATGATGTGTGTCAATGTGCTCAGAGCCACACCTTGTCTGTGACAGATGG 1329  
 Db 321 IalAlaProGlyProGlyGlyGly-----AlaAlaValProSerG 334  
 QY 1328 GGGCTGCTCTGAGAAAGTCTCTGCTCTGAGAGTTCTCATACACAGCTCTCTTCTGCT 1269  
 Db 334 IValA-----ThrProHisProGlnArgGlySerGlyP 345  
 QY 1268 CTGCA---CGACACATGAGAGATGCTGGCGCGCGGCTCTCAGAGAGGAAATCTGGCT 1212  
 Db 345 roAlaAspProProAlaAlaAlaArgLeuProProGlnArgGlnIleProArgLeuPro- 364  
 QY 1211 GGGCTTGAATGAGCAGAGAGCTGCTCTGTGCCAAAGTGGTCTGTGGTGGTGAAGCTG 1152  
 Db 365 -----GlnAspLeuAlaAlaGln----- 371  
 QY 1151 CTTTGCAGAGAAAGGGCTCTCAGCTTGCACAGCTGAGAGCTGAGAGTGGTTTA 1092  
 Db 371 ----- 371  
 QY 1091 GCGCAGATTCCTCCACGAACTGCGGTGCTGACAGGC-----GAGGGAG 1047  
 Db 372 -----ArgCysProAlaGlyProProProThrArgSerGly 384  
 QY 1046 GCGCGGCTTTCCGCGCGCGCGCGCTTCCAGGA-----CGAAGGCACTCCGACCG 993  
 Db 384 IalAlaIleAlaGlnArgThrHisArgArgProProGlyCysProArgSerAlaArgAsnProG 404  
 QY 992 GA-----CCCGGTC-----CGCGCGGAGCG-----GGGACGCGCCCTCCCGCG 951  
 Db 404 IYCySerProArgThrTrpArgArgArgSerGlyAlaGlnArgGlyHisProProPro- 422  
 QY 950 CGCGCGGGCGCGTGAAGGGGGGGCGCGCGCGCGGGGACAGCGCGGGGACCGGCTA 891  
 Db 423 -----GlyAlaGlyGlnArgProSerGlyProThrGlyAlaArgProAla 438  
 QY 890 TCCGAGGCCAACCGAGAGCTCCGCGCGCGGTATCGTTCGCTGCGCTGGGGGGAATTCTGA 831  
 Db 438 IalProGlyAlaPro-GlyThrProAlaAlaPro----- 448  
 QY 830 CTTAAGAGCGTTCACTCATTAATCCACAGATGTAAGTTCGCGCCATTTGGCTCTCAGCC 771  
 Db 449 -----GlyProGlyGlyAlaAla----- 454  
 QY 770 AAGCACAATACCAATATGTTGAACTGGGTTCTCTGCTGATGAGAGATTTCATG 711  
 Db 455 -----AlaValPro-----SerGly 459  
 QY 710 GCAACACACATCATCAGTAGGTAATACTAACTGTCTCAGCAGCGTCTAATCCAGCT 651  
 Db 460 AlalThrProHisProGlnArgGlySer----- 468  
 QY 650 CAAGTTCCCTATTAGTGGGTGAACAATCCAGCTTGGTAATTCGCTTCAATATGATA 591

```

Db 468 -----
QY 590 GGAAGAGCCGACATGAGAGATCAAAAAGAGAGCTGATGAGCGTTGGCCGACAA 531
Db 469 GYProAlaAsp-----ProProAla 475
QY 530 GCCAGTTATCCCTGTGTAACTTTCTGACACCTCTGCTTAAACCCAAAGGTCAAGA 471
Db 476 AlaAlaArg-----LeuProProGluArgGlnGlu 485
QY 470 GAGTGTGAGGCGCCGCTTACAGGCTGATGATGACTGAAATCAAGATCAAGCAGC 411
Db 486 -----ProArgLeuPro-----GlnAspLeuAlaAla 494
QY 410 TTT-----TGCCCTTCTGCTCCAGCGAGGTTTCTGCTCCCTGAGCTCCCTTAGCA 357
Db 495 AlaGlnArgCysProAlaGlyPro-----ProProThrArgSergly 508
QY 356 CACCTCGCTTACCGTTTGAACAGGTACCG-----CCCCAGTCAAACTCCCACT 306
Db 509 AlaAlaAlaGlnArgThrHisArgArgProProGlyCysProArgSeralArgAsnPro 528
QY 305 GGCAGTGTCCCGGAGCGGGTCCGCGCCGCGCGCGCGG-----CCGG 261
Db 529 Gly--CysProArgThrTyrArgArgArgSerglyAlaGlnArgGlyHisProProProG 548
QY 260 GCGCTTGGCGCCAGAACGAGAGACCCCTCGGGGCTGCCCC----- 220
Db 548 LysAlaGlyArgProSerglyProThrGlyGlyArgProAlaAlaProGlyAlaProG 568
QY 219 -----CCCGCTCAGCGGGTCACTGATAAAACGATAG--AGTAGTGATTTGACCGG 168
Db 568 LysThrProAlaAlaProGlyProGlyGlyAlaAlaAlaProSerglyAlaThrProH 588
QY 167 CGGCGCCGAGGCGCGCGAGACCCCGCGCGCGCTCCGCGGG--AAACCGGGGGGGG 111
Db 588 IsProGluArgGlySerglyProAlaAspProProAlaAlaAlaArgLeuProProGlu 608
QY 110 GCCCGGGGCGCCAGCTC 94
Db 608 TgGlnGluProArgLeu 613

RESULT 5
SRC8_MOUSE STANDARD; PRT; 546 AA.
AC 060598;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Src substrate cortactin.
GN EMS1 OR CTN.
OS Mus musculus (mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID:10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c;
RX MEDLINE=9426839; PubMed=7516062;
RA Mglarese M.R., Mannion-Henderson J., Wu H., Parsons J.T.,
RA Bender T.P.;
RT "The protein tyrosine kinase substrate cortactin is differentially
RT expressed in murine B lymphoid tumors.";
RL Oncogene 9:1989-1997 (1994).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 125-138, 273-289 AND 534-543.
RX MEDLINE=94043284; PubMed=7693700;
RA Zhan X., Hu X., Hampton B., Burgess W.H., Friesel R., Macias T.;
RT "Murine cortactin is phosphorylated in response to fibroblast growth
RT factor-1 on tyrosine residues late in the G1 phase of the BALB/c 3T3
RT cell cycle.";
RJ J. Biol. Chem. 268:24427-24431 (1993).

```

```

CC -1- FUNCTION: May contribute to the organization of cell structure.
CC The SH3 motif may function as a binding region to cytoskeleton.
CC Tyrosine phosphorylation in transformed cells may contribute to
CC cellular growth regulation and transformation.
CC -1- SUBUNIT: Interacts with SHANK2 and SHANK3 via its SH3 domain (By
CC similarity).
CC -1- TISSUE SPECIFICITY: Detected in most murine tissues, but not
CC detected in B lymphocytes or plasma cells.
CC -1- SIMILARITY: TO HEMATOPOIETIC LINEAGE CELL SPECIFIC PROTEIN.
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U03184; AAA19689.1; -.
CC PIR: I48899; I48899.
CC HSP: P19174; ZHSP.
CC MG: MG1:99695; Cten.
CC InterPro: IPR003134; Hal/Cortactin.
CC InterPro: IPR001018; Neu_cyt_fact_2.
CC InterPro: IPR01452; SH3.
CC Pfam: PF02218; HSI_rep_7.
CC Pfam: PF00018; SH3_1.
CC PRINTS: PR00499; P6PROX.
CC PRINTS: PR00452; SH3DOMAIN.
CC PRODOM: PD000066; SH3; 1.
CC SMART: SM00326; SH3; 1.
CC PROSITE: PS50002; SH3; 1.
CC Phosphorylation: Repeat: SH3 domain; Cytoskeleton.
CC KW DOMAIN 83 329 6.5 X 37 AA TANDEN REPEATS.
CC FT REPEAT 83 119 1.
CC FT REPEAT 120 156 2.
CC FT REPEAT 157 193 3.
CC FT REPEAT 194 230 4.
CC FT REPEAT 231 267 5.
CC FT REPEAT 268 304 6.
CC FT REPEAT 305 329 7 (INCOMPLETE).
CC FT DOMAIN 488 546 SH3.
CC FT CONFLICT 9 9 A -> R (IN REF. 2).
CC SQ SEQUENCE 546 AA; 61260 MW; 8F93A026AD1D6d4f CRC64;

Alignment Scores:
Pred. No.: 5,97e-06 Length: 546
Score: 231.00 Matches: 66
Percent Similarity: 47.80% Conservative: 21
Best Local Similarity: 36.26% Mismatches: 63
Query Match: 5,46t Indels: 32
DB: 1 Gaps: 8

US-10-028-952a-4 (1-2333) x SRC8_MOUSE (1-546)
QY 1026 GCGCGCGCGGAAAGCGCGCGCCCTCGCGCCGACGACCGACGTCGTGGGGAACC 1085
Db 395 AlaArgAlaLysGlnThrProPro----- 403
QY 1086 TGGCGCTAAACCACTTCATCTCCAGTCTTGAGCTGGCAAGCTGAGAGCCCTTCTTG 1145
Db 404 -----AlaSerProSerProGlnProLleGluAspArgProPro----- 416
QY 1146 CAGAGGAGCTCACCAACACAGAGACCCACTTGGCAGAGCGAGCTGCGCATCTCA 1205
Db 417 SerSerProLleTyrGluAspAlaAlaProPheLysAlaLysProSeryTyrArgGlySer 436
QY 1206 AGGCCAGGAGCATCTCCCTGCTGAGAGACCGCGCCGACGACCTCTCATGTGTG 1265
Db 437 GluProGluProGluTyrSerLleGlu-----AlaAlaGlyLleProGlu-----Ala 452
QY 1266 CAGCGAGAGAGAGCGGTGTGTAGAGAACTTCAGAGAGAGACCTCTAGAGACAG 1325

```











Db 2004 GlyIuserArglysalaleuGluValylsLysgluIuleuGlyAlaserSerProSer 2023  
 QY 475 -----GACCTTTGGTTTAAAGCAGAGGTCAGAAAAGTTACCAAGAGATTA 525  
 Db 2024 TyrGlyProProAlaLeuGlyPheValaSer-----ProSerSerGly 2038  
 QY 526 CTGGCTGTGGCGCCAGCCTTCATAGCGACGTGCTTTTGTATCCCTGCAGTCCGCT 585  
 Db 2039 ThrHisLeuGlyLeuGluLeuLysThrProaspValPheLysAlaProLeuThrPro 2058  
 QY 586 CTTCCTATCATTTGAGACAGAAATTCACCAAGCCTGGATTGTTCACCCACTATATAGGA 645  
 Db 2059 ArgAlaSerGluValGluProGlnSerProGlyLeuGlyLeu----- 2072  
 QY 646 ACGTAGCTGGATTAGACCGTCGTAGAGACAGGTATTTTACCTACTGATGATGTT 705  
 Db 2073 -----ArgPro-GlnGluPro-----ProProAlaGlnAlaLe 2083  
 QY 706 GTTGCCATGTATATCCCTGCTCAGTACGAGAGAACCGAGGTTGAGACTTGGTGTATG 765  
 Db 2083 uAlaPro----- 2085  
 QY 766 TGCTTGCTGAGAGCAATGGGGGAGAGTACATCTGTGAGATTATGACTGAACGCT 825  
 Db 2085 ----- 2085  
 QY 826 CTAAGTCAGATCCCGCCAGGCGAAGATACGACGAGCCCGCGAGCTCGGTGGCC 885  
 Db 2086 -----SerProProSerHisProAspIlePheArgProGlySerYrThrAspPr 2102  
 QY 886 TCGATACCGGATCCCGCCGCTGTCGCCGCGCGCGCGCGCGCGCGCGCGCGCGCG 931  
 Db 2102 o-TyrAlaGlnProProLeuThrProArgProGlnProProProProGlnuserGly 2121  
 QY 932 -----CCTCCACGCGCC-----CCGCGCGCGCGGAGGCGCGTCCCGCGCGCGCG 981  
 Db 2121 salaleuProProArgSerLeuProSerAspProPheSerArgValProValaSerProGln 2141  
 QY 982 GACCGCGGGTCGGGCGGAGAGCCCTCGTCCGAGAAAGGGGCGCGCGGAAAGGC 1041  
 Db 2141 nSerGlnSerSerSerGlnSerProLeuThrPro-----ArgProLeuSerAlaGluAl 2159  
 QY 1042 GCGCGCCCTCGCCCGCTCAGCGACCGCACCGTCTGTGG-----GAACCTGCGCGCTAAACCA 1098  
 Db 2159 aPheCysProSerProValThr--ProArgPheGlnSerProAspProTyrSerArgPro 2178  
 QY 1099 CCTCATCTCCAGT----- 1124  
 Db 2179 ProSerArgProGlnSerArgAspProPheAlaProLeuHisLysProProArgProGln 2198  
 QY 1125 AAGCTGAGAGAGCCCTTCCTGAGAGAGAGCTCAACCCACAGAGACCACTTGGAGAGA 1184  
 Db 2199 ProProGluValAlaPheLysAlaGlySerLeuAlaHisThrSerLeuGlyAlaGlyGly 2218  
 QY 1185 GAGCCAGCTGTGCTCATCTCAAGGCCCAAGGCAATCTCCCTGTGAGAGCCGCGCGCC 1244  
 Db 2219 PheProAlaAlaLeuProAlaGlyProAlaGlyLeuHisAlaLysValProSer-- 2237  
 QY 1245 AGCACTCTCCATGCTGTGAGAGAGAGAA----- 1274  
 Db 2238 GlyGlnProProAlaPheValArgSerProGlyThrGlyAlaPheValaGlyThrProSer 2257  
 QY 1275 -----GAGAGAGCTGTGATAG-----GAACCTCCAGAGAGAG 1307  
 Db 2258 ProMetArgPheThrPheProGlnAlaValaGlyGlnProSerLeuLysProProValPro 2277  
 QY 1308 GAGACCTTCTAGAGAGAGCCCACTGGTGCAGAGCAAGGTGTGCTGCAGACATTT 1367  
 Db 2278 Gln-ProGlyLeuProProProHisGlyTLeaSerHisPheGlyProGlyProThrIle 2297  
 QY 1368 G-----ACCAACCAATTTCAGAGGCGAGGGGCTC 1394  
 Db 2297 uGlyLysProGlnSerThrAsnYrThrValAlaThrGlyLysAsnPheHisProSerGlySe 2317

QY 1395 A--GTGGCAAGGGCTCT----- 1410  
 Db 2317 rProLeuGlyProSerSerGlySerThrGlyLysSerYrGlyLeuSerProLeuArgPr 2337  
 QY 1411 -----GTGCCGTTCCTGTACACTCCAGAGAGCGG-----AC 1445  
 Db 2337 oProSerValaLeuProProProAlaProAspIlePheArgProTyrLeuSerHisGlyAl 2357  
 QY 1446 GACACAGAGATCTCCCTTGACCCCGAGAACCTATCAACGCGCATCGAGGTATGACGAA 1505  
 Db 2357 aserGlnArgSerGlyIleThrSerProValaGluLysArg-----G1 2371  
 QY 1506 GGCTGTGCGCTGTGCTTANGGCGGATGCGCATTTTGGCATGTCCCTGCCAATGCG-- 1563  
 Db 2371 uAspProGlyThrGlyMetGly--SerSerLeuAlaThrAlaGluLeuProGlyThrG1 2390  
 QY 1564 -----TGAGGCTCATGTAGTGAGGCTGAGGCGACATCTTGCCCTCTCTCAG 1613  
 Db 2390 nAspProGlyMetSerGlyLeuSerGlnThrGluLeuGlyLysGlnArgGlnArgGlnArg 2410  
 QY 1614 ACATGCTTCCTTAT----- 1629  
 Db 2410 gLeuArgGluLeuLeuIleArgGlnGlnIleGlnArgAsnThrLeuArgGlnGluG1 2430  
 QY 1630 -----GCTGGAAGAGAGCGCTGGAGATTGAC----- 1656  
 Db 2430 uThrAlaAlaAlaAlaAlaGlyAlaValaGlyProProGlySerTrpGlyAlaGluProSe 2450  
 QY 1657 -----ATTACAGACTCTCCAGAAATAGACCC-----AGTAGAGATGA 1697  
 Db 2450 rSerProAlaPheGluGlnLeuSerArgGlyGlnThrProPheAlaGlyThrGlnAspY 2470  
 QY 1698 GGCCTCA-----GGCTCCCTCCGCTGCGAGTACAGCTGCTGACCCCAATGCGAGC 1751  
 Db 2470 sSerSerLeuValaGlyLeuProProSerSerLysSerGlyProIleLeuGlyProGlySe 2490  
 QY 1752 AATGCGCTGCTGATTCACACATCCCTCTCATCCCGGACCCCTCCAGAGACTTG 1810  
 Db 2490 r-PheProSerAspAspArgLeuSerArgProProProProAlaThrProSerSerMet 2509  
 Db  
 RESULT 9  
 ID 585 RAT STANDARD; PRT; 825 AA.  
 AC C63003;  
 DT 01-NOV-1997 (Rel. 35, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 01-NOV-1997 (Rel. 35, Last annotation update)  
 DE 585 antigen.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 RX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Water; TISSUE=Brain;  
 RA MEDLINE=96015159; PubMed=857300;  
 RA Suzuki E., Kojima N., Yoshimura K., Uemura K., Obata K., Akagawa K.;  
 RT "Cloning and sequence analysis of cDNA for a possible DNA-binding  
 protein 585 in the nervous system.";  
 RL J. Biochem. 118:122-128 (1995).  
 CC - FUNCTION: MIGHT HAVE DNA-BINDING ABILITY.  
 CC - SUBCELLULAR LOCATION: Nuclear.  
 CC - TISSUE SPECIFICITY: Expressed in neurons.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----











```

Db      563 asSerSerSerAlaSerAlaSerSerGlyValAlaValGlyAlaLeuGlyValArg----- 601
      |||
      ::|
      |||
Oy      1653 TGACATTCAGCACTCTTCCAGGAATAAGACCCCGAGAGATAGCCCTCAGAGCTCCC 1711
      |||
      |||
      |||
Db      602 -----GngJugIuThr-SerLeuGlyProArgAlaAlaSerGlyProArgGlyPro 619
      |||
      |||
      |||
Oy      1713 TCCGGCTTGCGACACTCAGCCTGTGACCCCAATGCA--CAATGCCCTGTGATTC 1765
      ::|
      ::|
      ::|
Db      619 rglYsCysAlaArgIuThrArgHLeaIuThrSerGlyAlaValProAlaGlyVal 639
      |||
      |||
      |||
Oy      1770 ACGCATTCCTTCGCACTCCCGGACCCCGCCAGACAGTGGGCTGGCCCTG----- 1822
      |||
      |||
      |||
Db      639 euhTrArgIuLeuProIleSerGlyAlaSerSerValAlaAlaLeuSerProGlyVal 659
      |||
      |||
      |||
Oy      1823 --GACAGATATCTAGCCCAAGCCCTGCGCTG 1852
      |||
      |||
      |||
Db      659 enlySThrIleThrGlyAerGlyCysLeuProIle 669
      |||
      |||
      |||

RESULT 12
CALL CHICK
ID CALL CHICK STANDARD; PRT; 1453 AA.
AC P02457;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-OCT-1989 (Rel. 12, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Collagen alpha 1(I) chain precursor.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinoptera; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP MEDLINE=8005315; PubMed=3678934;
RX Finner M.H., Boedtker H., Doty P.;
RA "Unusual DNA sequences located within the promoter region and the
RT first intron of the chicken pro-alpha 1(I) collagen gene.";
RL J. Biol. Chem. 262:13323-13332(1987).
RN [3]
RP MEDLINE=8007542; PubMed=2820966;
RX Finner M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
RA "Unusual DNA sequences located within the promoter region and the
RT first intron of the chicken pro-alpha 1(I) collagen gene.";
RL J. Biol. Chem. 262:13323-13332(1987).
RN [2]
RP MEDLINE=82231995; PubMed=7093229;
RX Kiang A.H., Gross J.; Corbett C., Dixit S.N., Yu W., Seyer J.M.,
RA Hargreaves J.H.;
RA "Amino acid sequence of chick skin collagen alpha 1(I)-C89 and the
RT complete primary structure of the helical portion of the chick skin
RT collagen alpha 1(I) chain.";
RL Biochemistry 21:2048-2055(1982).
RN [4]
RP MEDLINE=72243016; PubMed=5047697;
RX Eyre D.R., Glimcher M.J.;
RA "Evidence for a previously undetected sequence at the carboxy-terminus
RT of the alpha 1 chain of chicken bone collagen.";
RL Biochem. Biophys. Res. Commun. 48:720-726(1972).
RN [5]
RP MEDLINE=981-1453 FROM N.A.
RX MEDLINE=81160715; PubMed=6927845;
RA Fuller F., Boedtker H.;
RT "Sequence determination and analysis of the 3' region of chicken pro-
RT alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids
RL including the carboxy-terminal propeptide sequences.";
RL Biochemistry 20:996-1006(1981).
RN [6]
RP MEDLINE=82231995; PubMed=7093229;
RX Finner M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
RA "Unusual DNA sequences located within the promoter region and the
RT first intron of the chicken pro-alpha 1(I) collagen gene.";
RL J. Biol. Chem. 262:13323-13332(1987).
RN [3]
RP MEDLINE=8007542; PubMed=2820966;
RX Finner M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
RA "Unusual DNA sequences located within the promoter region and the
RT first intron of the chicken pro-alpha 1(I) collagen gene.";
RL J. Biol. Chem. 262:13323-13332(1987).
RN [2]
RP MEDLINE=82231995; PubMed=7093229;
RX Kiang A.H., Gross J.; Corbett C., Dixit S.N., Yu W., Seyer J.M.,
RA Hargreaves J.H.;
RA "Amino acid sequence of chick skin collagen alpha 1(I)-C89 and the
RT complete primary structure of the helical portion of the chick skin
RT collagen alpha 1(I) chain.";
RL Biochemistry 21:2048-2055(1982).
RN [4]
RP MEDLINE=72243016; PubMed=5047697;
RX Eyre D.R., Glimcher M.J.;
RA "Evidence for a previously undetected sequence at the carboxy-terminus
RT of the alpha 1 chain of chicken bone collagen.";
RL Biochem. Biophys. Res. Commun. 48:720-726(1972).
RN [5]
RP MEDLINE=981-1453 FROM N.A.
RX MEDLINE=81160715; PubMed=6927845;
RA Fuller F., Boedtker H.;
RT "Sequence determination and analysis of the 3' region of chicken pro-
RT alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids
RL including the carboxy-terminal propeptide sequences.";
RL Biochemistry 20:996-1006(1981).
RN [6]
RP MEDLINE=82231995; PubMed=7093229;
RX Finner M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
RA "Unusual DNA sequences located within the promoter region and the
RT first intron of the chicken pro-alpha 1(I) collagen gene.";
RL J. Biol. Chem. 262:13323-13332(1987).
RN [3]
RP MEDLINE=8007542; PubMed=2820966;
RX Finner M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
RA "Unusual DNA sequences located within the promoter region and the
RT first intron of the chicken pro-alpha 1(I) collagen gene.";
RL J. Biol. Chem. 262:13323-13332(1987).
RN [2]
RP MEDLINE=82231995; PubMed=7093229;
RX Kiang A.H., Gross J.; Corbett C., Dixit S.N., Yu W., Seyer J.M.,
RA Hargreaves J.H.;
RA "Amino acid sequence of chick skin collagen alpha 1(I)-C89 and the
RT complete primary structure of the helical portion of the chick skin
RT collagen alpha 1(I) chain.";
RL Biochemistry 21:2048-2055(1982).
RN [4]
RP MEDLINE=72243016; PubMed=5047697;
RX Eyre D.R., Glimcher M.J.;
RA "Evidence for a previously undetected sequence at the carboxy-terminus
RT of the alpha 1 chain of chicken bone collagen.";
RL Biochem. Biophys. Res. Commun. 48:720-726(1972).
RN [5]
RP MEDLINE=981-1453 FROM N.A.
RX MEDLINE=81160715; PubMed=6927845;
RA Fuller F., Boedtker H.;
RT "Sequence determination and analysis of the 3' region of chicken pro-
RT alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids
RL including the carboxy-terminal propeptide sequences.";
RL Biochemistry 20:996-1006(1981).
RN [6]
RP MEDLINE=82231995; PubMed=7093229;
RX Finner M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
RA "Unusual DNA sequences located within the promoter region and the
RT first intron of the chicken pro-alpha 1(I) collagen gene.";
RL J. Biol. Chem. 262:13323-13332(1987).
RN [3]
RP MEDLINE=8007542; PubMed=2820966;
RX Finner M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
RA "Unusual DNA sequences located within the promoter region and the
RT first intron of the chicken pro-alpha 1(I) collagen gene.";
RL J. Biol. Chem. 262:13323-13332(1987).
RN [2]
RP MEDLINE=82231995; PubMed=7093229;
RX Kiang A.H., Gross J.; Corbett C., Dixit S.N., Yu W., Seyer J.M.,
RA Hargreaves J.H.;
RA "Amino acid sequence of chick skin collagen alpha 1(I)-C89 and the
RT complete primary structure of the helical portion of the chick skin
RT collagen alpha 1(I) chain.";
RL Biochemistry 21:2048-2055(1982).
RN [4]
RP MEDLINE=72243016; PubMed=5047697;
RX Eyre D.R., Glimcher M.J.;
RA "Evidence for a previously undetected sequence at the carboxy-terminus
RT of the alpha 1 chain of chicken bone collagen.";
RL Biochem. Biophys. Res. Commun. 48:720-726(1972).
RN [5]
RP MEDLINE=981-1453 FROM N.A.
RX MEDLINE=81160715; PubMed=6927845;
RA Fuller F., Boedtker H.;
RT "Sequence determination and analysis of the 3' region of chicken pro-
RT alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids
RL including the carboxy-terminal propeptide sequences.";
RL Biochemistry 20:996-1006(1981).
RN [6]
RP MEDLINE=82231995; PubMed=7093229;
RX Finner M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
RA "Unusual DNA sequences located within the promoter region and the
RT first intron of the chicken pro-alpha 1(I) collagen gene.";
RL J. Biol. Chem. 262:13323-13332(1987).
RN [3]
RP MEDLINE=8007542; PubMed=2820966;
RX Finner M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
RA "Unusual DNA sequences located within the promoter region and the
RT first intron of the chicken pro-alpha 1(I) collagen gene.";
RL J. Biol. Chem. 262:13323-13332(1987).
RN [2]
RP MEDLINE=82231995; PubMed=7093229;
RX Kiang A.H., Gross J.; Corbett C., Dixit S.N., Yu W., Seyer J.M.,
RA Hargreaves J.H.;
RA "Amino acid sequence of chick skin collagen alpha 1(I)-C89 and the
RT complete primary structure of the helical portion of the chick skin
RT collagen alpha 1(I) chain.";
RL Biochemistry 21:2048-2055(1982).
RN [4]
RP MEDLINE=72243016; PubMed=5047697;
RX Eyre D.R., Glimcher M.J.;
RA "Evidence for a previously undetected sequence at the carboxy-terminus
RT of the alpha 1 chain of chicken bone collagen.";
RL Biochem. Biophys. Res. Commun. 48:720-726(1972).
RN [5]
RP MEDLINE=981-1453 FROM N.A.
RX MEDLINE=81160715; PubMed=6927845;
RA Fuller F., Boedtker H.;
RT "Sequence determination and analysis of the 3' region of chicken pro-
RT alpha 1(I) and pro-alpha 2(I) collagen messenger ribonucleic acids
RL including the carboxy-terminal propeptide sequences.";
RL Biochemistry 20:996-1006(1981).
RN [6]
RP MEDLINE=82231995; PubMed=7093229;
RX Finner M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
RA "Unusual DNA sequences located within the promoter region and the
RT first intron of the chicken pro-alpha 1(I) collagen gene.";
RL J. Biol. Chem. 262:13323-13332(1987).
RN [3]
RP MEDLINE=8007542; PubMed=2820966;
RX Finner M.H., Aho S., Gerstenfeld L.C., Boedtker H., Doty P.;
RA "Unusual DNA sequences located within the promoter region and the
RT first intron of the chicken pro-alpha 1(I) collagen gene.";
RL J. Biol. Chem. 262:13323-13332(1987).
RN [2]
RP MEDLINE=82231995; PubMed=7093229;
RX Kiang A.H., Gross J.; Corbett C., Dixit S.N., Yu W., Seyer J.M.,
RA Hargreaves J.H.;
RA "Amino acid sequence of chick skin collagen alpha 1(I)-C89 and the
RT complete primary structure of the helical portion of the chick skin
RT collagen alpha 1(I) chain.";
RL Biochemistry 21:2048-2055(1982).
RN [4]
RP MEDLINE=72243016; PubMed=5047697;
RX Eyre D.R., Glimcher M.J.;
RA "Evidence for a previously undetected sequence at the carboxy-terminus
RT of the alpha 1 chain of chicken bone collagen.";
RL Biochem. Biophys. Res. Commun. 4
```







QY	851	AAGCATACGGGACGCGCGGAGCGCTCGCTTGCGCTCGCATACCGAGTCCCGCGCGTTC	910
Db	144	SeirapserSerSerGlyLeuSerGlnGlyProAlaArgProGlyHisProProPheLeu	163
QY	911	CCCGCGCGGAGCGCGCGCGCGCGCTCCACGCGCGCGCGCGCGCGAGGAGCGCGTCC	967
Db	164	PheProProSerProProProProAlaPheProArgGlnProGlnSerGlyPheGln	183
QY	968	-----CCGCGCGCGCGC	979
Db	184	ProHisProSerValProProThnGlyTyHisAlaProMetGlnProProThnSerArg	203
QY	980	CGGAGCGGGGTCGCGTGGCGAGTGC-----CTTCCTCTGGGAAA	1021
Db	204	LeuPheGlnGlyProProProGlnAlaProProProHisProGlnLeuTyProGlySer	223
QY	1022	CGGGGCGCC-----GGCGGAAAAGCGGCGCGCGCGCGCGC	1056
Db	224	AlaGlyGlyValLeuSerGlyProProMetGlyProGlyGlyGlyAlaAlaAlaSer	243
QY	1057	CGTCACGACCGCAGCTTCGCGGGAACCTGGCGCTTAACCACTCCATCTCCAGTCTC	1116
Db	244	SerValGlyProProSerGlyGlyLys-----GlnHisProProProThnProLe	261
QY	1117	AGCGTGGAACTAGAGAGCGCTTCCTCGCGAGAGAGCTCACCCACACAGAGACCACT	1176
Db	262	ProLeuSerSerSerGlyAlaSerGly-AlaProProAlaLysPro-ProAsnThr	279
QY	1177	TTGGCAGAGACCGACCTGCTGCCATCTTCACAGGCCAGGCGAGATCTCCCTGCTGAGAGC	1236
Db	280	-----ProValGlyAlaGlyAsnLeuProSerAlaProProProAlaThrPhe	296
QY	1237	CG-----GCGCCGACGCTCTCCACAGTCTGCTGACGCGCAGAGAGAGGCTGTATG	1290
Db	296	roHisValThrProAsnLeuProPro-----	304
QY	1291	AGGAACCTCCAGAGCAGAGACCTTTCACAG-----CAAGCCCGACGTGTGAGC	1341
Db	305	-----ProProAlaLeuArgProLeuAsnAlaSerAlaSerProPro-----	319
QY	1342	AGCAAGGTGCTGCTCTGACGACATTGACACCACTTCAGAGGCGCAGGCGCTGAGGCGC	1401
Db	320	-----GlyMetGlyValaGlnProIleProGlyHisLeuProSerProHisAlaMetGlyG	338
QY	1402	AAGGCGCTGTGCGCGCGCTCTGACACACTACAGCAGCGCAGCAGACACAGATCT	1458
Db	338	InglyMetSerGly-LeuProProGlyProGlnGlySerGlyProThnLeuAlaProSerPro	357
QY	1459	-----CCTTTGACCGCCGAGACCTCATACAGGCGCATCTGAGATGATGACACAGAGGCTGATGCG	1515
Db	358	HisProLeuProProAlaSerSerSer-----AlaProGly	369
QY	1516	GTGCGTATAGGCGCGGATGCCATTTGGCATGTTCCCGCCAACTAGCTGAGCTCATG	1575
Db	370	ProProMet-AxTyTyProTySerSerCysSerSerSerSerSerValAlaHisSer	388
QY	1576	AGTGAAGCTGAGGAGACATCT-----TGCCCTTCCTCTTCAGA	1614
Db	389	SerSerSerAlaAlaThrSerGlnTyProAlaSerGlnThnLeuProSer-TyPro	407
QY	1615	CATGCGCTTCCTATATGCTGGAAGAGAGAGCGCGGAGTTGACATTCAGACACTCTCCAG	1674
Db	408	HisSerPhe-----	410
QY	1675	AATAGAGACCCCGAGTAGAGATGAGGCTTCAGGCGTCCCTCCGGCTTGCGAAGATCACT	1734
Db	411	---ProProProThnSerMetSerValSerAsnGlnProProGlyTyTy---ThrdInPro	428
QY	1735	GTACACCCCAAAATGACATGAGCTGGTGATTCACACACATCTCTCGATACCCCGAC	1794
Db	429	SerLeuProSerGlnAlaVal---Trp-----SerGlnGlyProProPro	442
QY	1795	CCTCCACAGACAGCTGGGCTCT--TGCCCTTCACAGATACTGAGCCAGC-----CCT	1845

Dd						1887
Oy	1846	GCCTGCGCCAGACCCCTGAATGGCACAATGCCAAGCTCGGGG----				
Dd	463	ProThrIgtlglYlnSerThrAlaHtSPROcLaLroAlaHtSHtSHtSGInGlN				482
Oy	1888	-----AAAGGCTTGAAGCAGGCGCATCTGGAGGCTTGCC				1923
Dd	483	GInGlInGlNPrOGInPrOGInPrOGInPrOGInGlNHtSHtSHtSGLYAnSergLY				501
 RESULT 15						
ID	W146_HUMAN	STANDARD;	PRT;	1336 AA.		
AC	OQ60JF8,					
DT	28-FEB-2003 (Rel. 41, Created)					
DT	28-FEB-2003 (Rel. 41, Last sequence update)					
DT	10-OCT-2003 (Rel. 42, Last annotation update)					
DE	WD-repeat protein WDCl46.					
OS	WDCl46.					
OS	Homo sapiens (Human).					
OC	Eumetazoa; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;					
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
OX	NCB1_TaxId=9606;					
RN	[1]					
RP	SEQUENCE FROM N.A., SUBCELLULAR LOCATION, AND TISSUE SPECIFICITY.					
RA	MEDLINE=21092660, PubMed=11162572;					
RA	Ito S., Sakai A., Nomura T., Miki Y., Ouchida M., Sasaki J.,					
RA	Shimizu K.;					
RT	"A novel WD40 repeat protein, WDCl46, highly expressed during					
RL	spermatogenesis in a stage-specific manner.";					
RL	Biochem. Biophys. Res. Commun. 280:656-663(2001).					
CC	-1- SUBCELLULAR LOCATION: Nuclear.					
CC	-1- TISSUE SPECIFICITY: Most highly expressed in testis.					
CC	-1- SIMILARITY: Contains 1 collagenous domain.					
CC	-1- SIMILARITY: Contains 7 WD repeats.					
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration					
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC	the European Bioinformatics Institute. There are no restrictions on its					
CC	use by non-profit institutions as long as its content is in no way					
CC	modified and this statement is not removed. Usage by and for commercial					
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/					
CC	or send an email to license@sib-ch).					
DR	EMBL; AB044749; BAB32435.1;-					
DR	GO; GO:0005634; C:nucleus; IDA.					
DR	GO; GO:0006301; P:postreplication repair; NAS.					
DR	GO; GO:0007283; P:spermatogenesis; NAS.					
DR	InterPro; IPR008160; Collagen.					
DR	InterPro; IPR001680; WD40.					
DR	pfam; PF001391; Collagen; 1.					
DR	pfam; PF004600; WD40; 7.					
DR	ProDom; PD000018; WD40; 2.					
DR	SMART; SM00320; WD40; 7.					
DR	PROSITE; PS00678; WD_REPEATS_1; FALSE_NEG.					
DR	PROSITE; PS00082; WD_REPEATS_2; 6.					
DR	PROSITE; PS00294; WD_REPEATS_REGION; 1.					
KM	Collagen; Repeat; WD repeat; Nuclear protein.					
FT	REPEAT	117	WD 1.			
FT	REPEAT	159	WD 2.			
FT	REPEAT	198	WD 1.			
FT	REPEAT	200	WD 3.			
FT	REPEAT	242	WD 4.			
FT	REPEAT	286	WD 5.			
FT	REPEAT	329	WD 6.			
FT	REPEAT	373	WD 7.			
FT	DOMAIN	6-8	COLLAGEN-LIKE.			
SQ	SEQUENCE	1336 AA;	145921 MW;	865E98A2D12381F6 CRC64;		
 Alignment Scores:						
pred. No.:	0.00251	Length:	1336			
Score:	164.50	Matches:	167			
Percent Similarity:	28.65%	Conservative:	43			





Fri Apr 2 07:35:04 2004

us-10-028-952a-4.rsp

Page 22

```
QY 142 CCCCCGCCCCCTGCGGGGACACCGGGGGGGCC----- 108
Db 1300 erglythrProSerArglyGlyArgSerGlySerAentrpGlyArgGlySerAsnMetA 1320
QY 107 --GGGGGCCCAAGCTCAACAGGGTCTCT 81
Db 1320 snserGlyProProArgArgGlyAlaSer 1329
```

Search completed: April 1, 2004, 15:54:43  
Job time : 77 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comphen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 1, 2004, 15:49:01 ; Search time 40 Seconds  
(without alignments) 11220.742 Million cell updates/sec

Title: US-10-028-952A-4  
Perfect score: 4230  
Sequence: 1 gacagagatccactgtcc.....agcttgactcttcgcca 2333

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 566732

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ n2p model -DRV=xlp  
-O=/cgn2.1/USFEO.spool\_p/US10028952/runat.01042004.154626.670/app.query.fasta.1.2503  
-DB=PIR\_78 -QPM=fastan -SUFFIX=rpr -MINMATCH=0.1 -LOOPEXT=0  
-UNITS=bites -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45  
-DOCCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US10028952 @CEN 1.1 60 @unat 01042004.154626.670 -NCPU=6 -ICPU=3  
-NO MAP -LARGEJOINT -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGCLOG  
-DRV TIMEOUT=120 -MAP TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  
PIR\_78:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	288.5	6.8	643	2 T23453	hypothetical prote
2	277	6.5	486	2 S07633	hematopoietic line
3	261.5	6.2	486	2 A49760	lckbpl protein - m
4	244.5	5.8	563	2 A41530	src substrate p80/
5	243	5.7	270	2 T02955	probable cytochrom
6	242.3	5.7	660	1 Q08B3	BHLF1 protein - hu
7	231	5.5	546	2 I48899	coractctn - mu
8	230	5.4	550	2 A48063	mammary tumor/squa
9	215.5	5.1	1106	2 J00405	hypothetical 119.5
10	199	4.7	530	2 T02995	unspecific monocly
11	195.5	4.6	4957	2 T03455	ALR protein - huma
12	195.5	4.6	5262	2 T03454	ALR protein - huma
13	193.5	4.6	825	2 J04163	DNA-binding protei
14	191.5	4.5	660	1 Q08B3	BHLF1 protein - hu

15	191	4.5	775	1 EDBE11	immediate-early pr
16	189.5	4.5	1414	1 S23809	collagen alpha 2(I)
17	189	4.5	1042	1 CGCH1S	collagen alpha 1(I)
18	188	4.4	167	2 H72579	hypothetical prote
19	185	4.4	862	2 T46289	hypothetical prote
20	185	4.4	1105	2 J00405	hypothetical 119.5
21	184.5	4.4	1453	2 S21626	collagen alpha 1(I)
22	182	4.3	592	1 ILB1	actin-binding prot
23	180.5	4.3	315	2 A56561	35K proline-rich p
24	180	4.3	617	2 S42719	actin-binding prot
25	179.5	4.2	1497	2 I49607	procollagen type V
26	178.5	4.2	1373	1 A43291	collagen alpha 2(I)
27	178	4.2	390	2 G01936	ABL binding protei
28	178	4.2	699	2 C43674	US4 protein - huma
29	177.5	4.2	1464	2 S59856	collagen alpha 1(I)
30	176	4.2	1838	1 CGHUV	collagen alpha 1(I)
31	175.5	4.1	1215	2 T32734	myosin-Ib - Acanth
32	175.5	4.2	1466	1 CGHUTL	hypothetical alpha 1(I)
33	175.5	4.2	1585	2 T31611	hypothetical prote
34	175.5	4.1	2715	2 T13049	eyelid - fruit fly
35	175	4.1	290	2 T42526	hypothetical prote
36	175	4.1	1184	2 G01763	atrophin-1 - human
37	175	4.1	3530	2 A59286	unconventional myo
38	174	4.1	903	2 T00705	N-chimerin homolog
39	174	4.1	1419	2 A41182	collagen alpha 1(I)
40	174	4.1	1487	2 B41182	collagen alpha 1(I)
41	173.5	4.1	1366	1 CGHUS	collagen alpha 2(I)
42	172.5	4.1	1464	1 CGHUS	collagen alpha 1(I)
43	172	4.1	416	1 SKX1AG	dermal gland prote
44	172	4.1	1151	2 T18535	high molecular mas
45	172	4.1	1184	2 S50832	atrophin-1 - human

## ALIGNMENTS

RESULT 1  
T23453  
Hypothetical protein K08B3.4 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
R:McMurray, A.  
Submitted to the EMBL Data Library, November 1996  
A:Reference number: Z19743  
A:Accession: T23453  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-643 <WIL>  
A:Cross-references: EMBL:Z81568; PIDN:CA804592.1; GSPDB:GN00021; CESP:K08B3.4  
A:Experimental source: clone K08B3  
C:Genetics:  
A:Gene: CESP:K08B3.4  
A:Map position: 3  
A:Introns: 47/1; 84/3; 185/2; 255/2; 455/1

## Alignment Scores:

Pred. No.: 7.98e-12 Length: 643  
Score: 268.50 Matches: 92  
Percent Similarity: 38.04% Conservative: 32  
Best Local Similarity: 28.22% Mismatches: 105  
Query Match: 6.82% Indels: 97  
DB: 2 Gaps: 11

US-10-028-952A-4 (1-2333) x T23453 (1-643)  
QY GCCCAGCGGGAACGATAGCGAGCGCGGAGCCTCGTTCGCTCGGATAGCGGCTTC 900  
DB 335 AAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 354  
QY CCCGCGCTGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGAGG 960  
DB 355 LysAlaGlnIleSerAspIlyProValProSerSerValaAnProAla----- 370

**RESULT 2**

S07633  
hematopoietic lineage cell-specific protein HSI - human  
C:Species: Homo sapiens (man)  
C>Date: 29-Jan-1993 #sequence revision 29-Jan-1993 #text change 05-Nov-1999  
C:/Date: S07633; A47478; B47478; C47478; D47478; E47478  
R/Kitamura, D.; Kaneko, H.; Miyagoe, Y.; Ariyasu, T.; Matanabe, T.  
Nucleic Acids Res. 17, 9367-9379, 1989  
A>Title: Isolation and characterization of a novel human gene expressed specifically in  
A:Reference number: S07633; PMID:2587259  
A:Accession: S07633  
A:Molecule type: mRNA  
A:Residues: 1-466 <KIT>

Dy CGCGTGGCCGCGCGCGCGCGAGACCGGAGTCGGATGCCCTTGCTGTGGAA 1020  
Dy 371 -----ProAlaProAlaPheThrProSerAlaIleLeu---ValGlySerArgLysGlu 387  
Qy 1021 ACCGGGGCGCGCGCGAAAAGCGCGCGCCCCCTGCCTCCGTCACGACCGCACATTGTTGGG 1080  
Db LeuPheSerSerLysProSerGlyProValLeuProLys-----ProGlnValAsnGly 405  
Qy 1081 GAACCTGCGCTTAACAACCTCCATCTCCACTCTCAAGCTTAGCGAGCTGAAGAGCCCTT 1140  
Db 406 SerProLysLysTrpProPro--ValGlyInrThrSerProPro----ArgGluProV 423  
Qy 1141 TCCTGCAGAACAGCTCACCCACCAGAG-----ACCCAATTGG 1179  
Db 423 al-----AsnArgLeuThrGluProAspGluProSerThrTyrrThroLysProIlea 441  
Qy 1180 GCAGAAGACCCAGCTGCTGCCCATCTTCAGAGCCCAGGGAGATCTCCCTGCT----- 1229  
Db 441 lAtyrGluProGluProMetValTyrLysProGluAlaMetLysProAlaValSerTyrA 461  
Qy 1230 -----CAGAGACCGGCGCCACGACT----- 1250  
Db 461 spAlaTyrGluGluProProAlaGluProAlaProProThrPheLeuAlaProThrProV 481  
Qy 1251 -----CCTCCATGTCTGTGTCACGACGACGAGAGAGAGGCTGTGNTG 1290  
Db 481 AlIleAlaProProProProGluProThrProAlaProSerHisTyrAlaSerGlnTyrA 501  
Qy 1291 AGGAACCTCCAGAGAGAGACCTTC----- 1316  
Db 501 spAlaProProValHisGluSerPheGluProAlaValProProValSerAlaProSerH 521  
Qy 1317 -----TACGAGCAGCCCCCACTGTGTGACAGACAGACGAGTGTGCTGTGAGC 1362  
Db 521 lStYralaSerGlnTyrAspAlaProProGluProLaspserHisSerSerSerg 541  
Qy 1363 ACATTGACCACCACTTCAGGGCCAG----- 1388  
Db 541 InLeuProAlaHisIleAlaSerGlnTyrAspMetProProValMetProGluInProV 561  
Qy 1388 ----- 1388  
Db 561 alphaAlaProLysSerSerProIleLysValAlaAlaProProIleLaspGlnTyrAsp 581  
Qy 1389 -----GGGCTCAGTGGGCGAAGGCTGTGTGCCCCGCTGTGACACTTCACGAGCAG 1440  
Db 581 heProProAlaValAlaGluGlnAsn-----AlaMetAlaLeuThrAspTYrgInAlaA 599  
Qy 1441 CCAGAGACACAGATTCCTTGAACCCGAGAACCTCATCAGCGAGCATCGAGTATCG 1500  
Db 599 laasphasphurJulleserPheAspProAspAspIleIethrasplileaspGlnValA 619  
Qy 1501 AGGAAGCTGTGTCGCGTGTATGAGCCGAGATGCCATTGTGGCATGTTCCCTGCCACT 1560  
Db 619 sSpSerGlyTyrPrtpylsglyArgAlaProSerGlyArGValGlyLeuPheProAlaAsnt 639  
Qy 1561 ACOTGAGACTCAT 1574  
Db 639 yrvallylsLeulle 643

A:Cross-references: EMBL:X16663; NID:G32054; PDB:G32055  
A:Yamanashi, Y.; Okada, M.; Semba, T.; Yamori, T.; Umemori, H.; Tsunawawa, S.; Toyoshima  
Proc. Natl. Acad. Sci. U.S.A. 90, 3631-3635, 1993  
A:Title: Identification of Hs1 protein as a major substrate of protein-tyrosine kinase (t  
A:Reference number: A47478; MUID:93234551; PMID:7682714  
A:Accession: A47478  
A:Status: preliminary  
A:Molecule type: protein  
A:Residues: 4-19, 'XXT', 23-26, 'X', 79-93, 'X', 95, 'X', 134-146, 208-223, 274-284, 'X', 286, 'X', 2  
A:Experimental source: Daudi, B-lymphoblastoid cells  
A:Note: sequence modified after extraction from NCI backbone  
C:Superfamily: SH3 homology  
F:435-482/Domain: SH3 homology <SH3>

Alignment Scores:  
Pred. No.: 4.77e-11 Length: 486  
Score: 277.00 Matches: 87  
Percent Similarity: 42.97% Conservative: 23  
Best Local Similarity: 33.98% Mismatches: 85  
Query Match: 6.55% Indels: 61  
DB: 2 Gaps: 10

US-10-028-952A-4 (1-2333) x S07633 (1-486)

QY	938	CGCCCCCGCGCGCGCGGAGGAGGCGGCGTGC	CGCCCGCGCGCGCGG	-----GACCGGGG	991
Db	250	ArglysaArglugluglugluglysalaglnvalalalarghglnglnglnlarys			269
QY	992	CCGGTG-----CGAGTGCCTTCGCTCGGAAACGGGGCGCGCGGAAAGCGCC			1045
Db	270	AlalvalThrlysaRgsSerProglualProglinProvallIlealmeuglugluproala			289
QY	1046	GCCCCCTCCGCCGCTCAGCACCGGACGCTGCGGGAAACCTGGGCTAAACCTCAT			1105
Db	290	ValProalalProleuProluyls			297
QY	1106	CTCCAGTCCCTCAGCCTGGCAGCTGAGAGCCCTTCCTCGAAGACGCTACCCAAC			1165
Db	298	IleerSerglualatProProvalglYThrProProSerSerSerSerSergluProval			317
QY	1166	AGAGACCCACTTTGGCAGAGAGCCAGCTGCATCTCAAGGCCCGAGGAGATTTGCC			1225
Db	318	Arg-Thr-----SerargluHlsProvalProleuProlleatrglnThrleupr			335
QY	1226	TGCT-----GAGAGCGCGCGCCAGACACTCTCCATGTCTG-----GTGCGAGC			1270
Db	335	CGluaspasnugluPro-----ProalaleuProProkgrThreugluglyleuglnva			354
QY	1271	AGAAAGAGAGGCTGTGTATGAGAACTCCAGAGCAGAGACCTTCTACGAGCAGCCCC			1330
Db	354	IGlulgugluProvalTyrgluHlsagluProgluProgluPro-----GluProgluPr			372
QY	1331	ACTGGTGAGAGCAGAGGTGCTGGCTGAGACATTGACACCATTTAG-----			1382
Db	372	CGluProgluasnleprTyrgluaspvalglugluMetasparghlsGluuglngluaspgl			392
QY	1382	-----			1382
Db	392	uProgluglyasprTyrglugluvalleugluProgluaspSerSerPheSerSerAlale			412
QY	1383	-----GGCAGGGGGCTCACTGGGAGGGCTCTG			1411
Db	412	uAlaglySerSerglyCysProalaglyAlalaglyAlalaglyAlalaleuglyIlele			432
QY	1412	TGCCCGTCCCTGTACGACTACAGAGCAGCCGACAGACACAGAGATCTCTTGACCCGA			1471
Db	432	rlalvalalaleuTyraapTyrglnglygluglySeraspgluSerPheaspProas			452
QY	1472	GAACTCATCAGGGCATCGAGGTATCGAGAAAGCTGGCGGTGGCTAAGGGCGGA			1531
Db	452	pasvallelThraspIlegluMetvalaspIuglyTyprtpargly--ArgCysHl			471
QY	1532	TGGCATTGTGGCATGTCCCTGCACCACTACGTGAGCTCATTTAG			1577

Db 471 sglYhSphegLYLeuPheProAlaSnTYrValLYsLeuGlu 486

## RESULT 3

149760  
ICKB1 protein - mouse

N/Alternate names: HSI

C/Species: Mus musculus (house mouse)

C/Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 05-Nov-1999

C/Accession: I49760; S57244; S52474

R/Kitamura, D.; Kaneko, H.; Taniuchi, I.; Akagi, K.; Yamamura, K.; Watanabe, T.

Biochem. Biophys. Res. Commun. 208, 1137-1146, 1995

A/Title: Molecular cloning and characterization of mouse HSI.

A/Reference number: I49760; MUID:95217198; PMID:7535527

A/Accession: I49760

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-486 <RES>

R/Takenoto, Y.; Furuta, M.; Li, X.K.; Strong-Sparks, W.J.; Hashimoto, Y.

EMBO J. 14, 3403-3414, 1995

A/Title: LCKB1, a proline-rich protein expressed in haematopoietic lineage cells, direct

A/Reference number: S57244; MUID:95354658; PMID:7628441

A/Status: preliminary

A/Accession: S57244

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-486 <TRK>

A/Cross-references: EMBL:X84797; NID:G683480; PIDN:CA55265.1; PID:G683481

C/Superfamily: SH3 homology

F/436-483/Domain: SH3 homology <SH3>

## Alignment Scores:

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
5.37e-10	261.50	486	92	22	107	77	12
Percent Similarity:	38.26%						
Best Local Similarity:	30.87%						
Query Match:	6.18%						

US-10-028-952A-4 (1-2333) x I49760 (1-486)

```

QY 777 GGAAGCCATGGG-----GGGAAGCTACCATCTGTGGATTATGACGAACGCTCTAG 830
Db 234 GYAlaArgGlyLeuLYsAlaLYsPheGluSerLeuLactGluLYsArgLYsArgLYs 253
QY 831 TCAGATCCCGCCCGGAGCGAGATACGACGACGCG-----CGCGAGCTTGATGGC 884
Db 254 GAGGATCCCGCCCGGAGCGAGATACGACGACGCG-----CGCGAGCTTGATGGC 273
QY 885 CTGGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 938
Db 274 MetSerArgGluValGlnGlnProSerMetProValGlnGluProAlaAlaProAlaGln 293
QY 939 GCGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 998
Db 294 LeuProLYsLYsLYsSerSerGluValTrpProPro----- 305
QY 999 GGAAGCTCTTCTGCTGCTGGAAGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1058
Db 306 -----AlaGluSerHisLeuProProGluSerGlnPro 316
QY 1059 TCAGCAGCGGAGCTGCTGCTGGAAGCTGAGCGCTAAACACATCTCATCTCTGAG 1118
Db 317 ValArgSerArgArgGlyTr-----ProVal 325
QY 1119 CTTGGC---AAGCTGAGAGCGCGCTTCTGCGAAGAGCTCACCAACGAGAGACCCAC 1175
Db 326 ProSerLeuProThrArgGlnSerProLeuGlnAsnHisLeuGluAspAsnGlu----- 343
QY 1176 TTGGAGAGAGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1232
Db 344 -----GluProProAlaLeuProProArgThrProGluGluValValGlu 360

```

QY 1223 GAGCGGCGCGCGCGAGCTCTCCATGTCGTGAGCGAGAGAGAGCTGTATAG 1292

Db 361 GlnProValLYrGluAlaAlaProGluLeu-----GluProGluProGluProAspTYrGlu 379

QY 1293 GAACTCCAGAGCAGAGAGACCTTCTACAGCAG----- 1325

Db 380 ProGluProGluThrGlnProAspTYrGluAspValGlyGluLeuAspArgGlnAspGlu 399

QY 1326 -----CCCACTGCTGAG 1340

Db 400 AspAlaGluGlyAspTYrGluAspValLeuGluProGluAspThrProSerLeuSerTYr 419

QY 1341 CAGCAGAGCTGCTGCTGCTGAGCATTGACCACTTACCACTTACAGGCGCGAGCTGAG 1400

Db 420 GlnAlaGlyProSerAla-----GlyAlaGlyGly 429

QY 1401 CAAGGCTCTTGTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1460

Db 430 AlaGlyLeuSerAlaLeuAlaLeuLYsAspTYrGlnGlyGlySerAspGluLeuSer 449

QY 1461 TTGGACCCGAGAACCTCATCATCGAGCGGATCGAGGTGATCGAGAGGCTGTGCGTGC 1520

Db 450 PheAspProAspAspIleIleThrAspIleGluMetValAspGluGlyTrpTrpArgGly 469

QY 1521 TATGGCGCGATGGCCATTGTCATGTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1574

Db 470 Gln--CysArgGlyHisPheGlyLeuPheProAlaSnTYrValLYsLeuLeu 486

QY 1574

Db 486

QY 1574

Db 486

QY 1574

Db 486

QY 1574

Db 486

QY 1574

Db 486

QY 1574

Db 486

QY 1574

Db 486

QY 1574

Db 486

QY 1574

Db 486

QY 1574

Db 486

QY 1574

Db 486

QY 1574

Db 486

QY 1574

Db 486

QY 1574

Db 486

QY 1574

Db 486

QY 1574

Db 486

QY 1574

Db 486



Db 185 rGPrOAlaAlaProGlyAlaProGlyThrProAlaAlaProGlyProGlyGlyAla 205  
 QY 1568 ATGTAACCTCCAGGCTCTCTCCAGCAATAGAGCCATGTCTGAGAGGGAGG 1599  
 Db 205 lAvalProSerGlyAlaThrProHisProGluArgSerGlyProAlaAspProPro 225  
 QY 1598 CAAATGTGCTTACGCTCACTCAATAGAGCTCCAGTGTGGCAGGGA 1549  
 Db 225 lAAlaAlaArgLeuProProGluArgGlnGluProArgLeuProGlnAspLeuAla 245  
 QY 1548 -----ACATGCCAAATAGG-----CATCGGGCCATAGCAAGCCACGACTT----- 1504  
 Db 245 lAGlnArgCysProAlaGlyProProProThrArgSerGlyAlaAlaGlnArgThn 265  
 QY 1503 --CGTCATCACTCGATGCCCGGTGATGAGGTTCTCGGGGTCAAAGAGATCTGTGT- 1447  
 Db 265 lAArgArgProProGlyCysPro-----ArgSerAlaArgAsnProGlyCysP 281  
 QY 1446 -----CGTCGGCTGCTGTAGTCTGACAGGGCAG----- 1417  
 Db 281 rOAlaGlnThrPArgArgSerGlyAlaGlnArgGlyAlaProProProGlyAlaGly 301  
 QY 1416 -----GGGACAGAGCCCTTGGCCCATGAGCCCTGAG----- 1384  
 Db 301 lNArgProSerGlyProThrGlyGlyAlaProAlaAlaProGlyAlaProGlyThrPro 321  
 QY 1383 -----CCTGAATGTGTGTCAATGTGTCTGAGAGCCAGCACTTGTGTCTGACCACTG 1329  
 Db 321 lAAlaProGlyProGlyGlyAla-----AlaAlaValProSerG 334  
 QY 1328 GAGGTCTGTGAGAGGTTCTCTCTGAGAGTTCTCATACAGACTCTCTTCTG 1269  
 Db 334 lYAla-----ThrProHisProGluArgGlySerGlyP 345  
 QY 1268 CTGCA-----CCAGACATGAGAGAGTGTGGGCGCGGCTCTCTGACAGAGAGATGTCCT 1212  
 Db 345 rOAlaAspProProAlaAlaAlaArgLeuProProGluArgGlnGluProArgLeuPro 364  
 QY 1211 GGGCTTGAAGTGAAGTGTGCTCTGACGCAAGTGGGTCTGTGGTGGAGAGCTG 1152  
 Db 365 -----GlnAspLeuAlaAlaGln----- 371  
 QY 1151 CTTCTGAGAGAGGGCTCTCTGAGCTTCCAGGCTGAGAGTGGAGGTGTTA 1092  
 Db 371 ----- 371  
 QY 1091 GCGCCAGAGTCCCAAGAGAGTGGCGTGAAGGCG-----GAGGGG 1047  
 Db 372 -----ArgCysProAlaGlyProProProThrArgSerGlyAla 384  
 QY 1046 GCGCGCTTTCGCGCGCGCGCTTCCAGAG-----CGAAGGCACTCCGACCG 993  
 Db 384 lAAlaAlaGlnArgThnHisArgArgProProGlyCysProArgSerAlaArgAsnProG 404  
 QY 992 GA-----CCCCGGTCC-----CGGCGCGCGCG-----GGGCGCGCGCGCTCCCGG 951  
 Db 404 lYCysProArgGlnThrPArgArgArgSerGlyAlaGlnArgGlyAlaProPro----- 422  
 QY 950 GCGCGCGCGCGCTGAGAGGGGGCGCGCGCGCGCGCGAGACAGCGGGGGAGCCGCTA 891  
 Db 423 -----GlyAlaGlyGlnArgProSerGlyProThrGlyGlyArgProAla 438  
 QY 890 TCCGAGGCGCAACGAGGCTCCGCGCGCTCCGATGATGTTCCGCTGGGGGAGATTCTGA 831  
 Db 438 lAProGlyAlaPro-GlyThrProAlaAlaPro----- 448  
 QY 830 CTTAGAGCGCTTACGTATATCCCAAGATGTAGATTGCCCCATTTGCTCTCAGCC 771  
 Db 449 ----- 454  
 QY 770 AAGCACAACCAATGTCTGAACCTGCGCTCTCTCTCTACTGAGCAGAGATTACATG 711  
 Db 455 -----AlaValPro-----SerGly 459

QY 710 GCACACACATCATAGTAGGTTAAACTAATCACTGTCTCAGAGCGTCTAATCCAGCT 651  
 Db 460 AlAthrProHisProGluArgGlySer----- 468  
 QY 650 CACGTTCCCTATTAGTGGGTGAACAATCAACGCTTGGTGAATTTGCTTCAATGATA 591  
 Db 468 ----- 468  
 QY 590 GGAAGAGCGACATGAGATCAAAAAGCAGCTGCTATGAACGCTTGGCGGCACAA 531  
 Db 469 GlyProAlaAsp-----ProProAla 475  
 QY 530 GCGAGTTATCCCTGTGTAATTTTGTACACCTCTCTTAAACCAAAAGTCAAGAA 471  
 Db 476 AlAAlaArg-----LeuProProGluArgGlnGlu 485  
 QY 470 GGAATGTAGAGCCCGCTTACAGGCTGTATTTGTAAGTGAATAATCAAGATCAAGCAGC 411  
 Db 486 -----ProArgLeuPro-----GlnAspLeuAlaAla 494  
 QY 410 TTT-----TGCCTTCTGTCTCAGAGAGGTTTGTCTCTGAGCTGCTTGAAGA 357  
 Db 495 AlAGlnArgCysProAlaGlyPro-----ProProThrArgSerGly 508  
 QY 356 CACCTGCGTTACCGTTTGAACAGTGTACG-----CCCACTCAACTCCCACT 306  
 Db 509 AlAAlaGlnArgThnHisArgArgProProGlyCysProArgSerAlaArgAsnPro 528  
 QY 305 GGCAGTGTCCCGAGAGCGGATGCGCGCGCGCGCGCG-----CCG 261  
 Db 529 Gly--CysProArgGlnThrPArgArgArgSerGlyAlaGlnArgGlyHisProProG 548  
 QY 260 GCGCTTGGCGCGCAAGAGGAGAGCCCTCGGGGCTCGCC----- 220  
 Db 548 lYAlaGlyGlnArgProSerGlyProThrGlyGlyArgProAlaAlaArgLeuProGlu 568  
 QY 219 -----CCGCGCTCACCGGCTGAGTGAATAAAGATCAG-----AGTAGTGTATTCACCGG 168  
 Db 568 lYThrProAlaAlaProGlyProGlyGlyAlaAlaValProSerGlyAlaThrProH 588  
 QY 167 CGGCGCGAGGGCGCGCGAGCCCGCGCGCGCGCTTGGCGG-----ACACCGGGGGGCG 111  
 Db 588 lSProGluArgGlySerGlyProAlaAspProProAlaAlaAlaArgLeuProGlu 608  
 QY 110 GCGGGGGGCGCAAGCTC 94  
 Db 608 rGlnGlnProArgLeu 613

RESULT 7  
 14889  
 cortactin - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999  
 C:Accession: I48899; A49514  
 R:Magdarsse, W.R.; Wamton-Henderson, J.; Wu, K.; Parsons, J.T.; Bender, T.P.  
 Oncogene 9, 1989-1997, 1994  
 A:Title: The protein tyrosine kinase substrate cortactin is differentially expressed in:  
 A:Reference number: I48899; PMID:94268839; PMID:7516062  
 A:Accession: I48899  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1566 <RBS>  
 A:Cross-references: EMBL:U03184; NID:9414990; PIDN:AAA19689.1; PID:9509495  
 R:Chan, X.; Hu, X.; Hampton, B.; Burgess, W.H.; Friesel, R.; Maciag, T.  
 J. Biol. Chem. 268, 24427-24431, 1993  
 A:Title: Murine cortactin is phosphorylated in response to fibroblast growth factor-1 on  
 A:Reference number: A49514; PMID:94043284; PMID:7693700  
 A:Accession: A49514  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid; protein  
 A:Residues: 1-8, 'R', 10-297, 'S', 299-546 <ZNA>  
 A:Experimental source: BALB/C 3T3 cells







```

Db      703 HistHglYalagYala-AlaserArgPro----- 712
Qy      489 AAAACCCAAAGTCAGAGAGATGCTGAGCGCCGCTTTCACGCTGTATGCTACTGA 430
Db      713 -----GlyGluGlyAla-----ArgAlaAs 719
Qy      429 AAATCAAGATCAAGGAGACTTTGCTCTCTGCTCAACGAGAGTTTCTGTCTCTCTGA 370
Db      719 p-----GlyProArgArgProGlyGluGlnProGl 729
Qy      369 GCTCCCTTTAGGA-----CACTGCTTACCTTT 340
Db      729 uGlyArgLeuGlyProGlyProAlaArgGlyProHleGlyArgAspGlyValArgyle 749
Qy      339 GACAGGTGTACCGCCCGCAGTCAAACTCCCGACCTGCTCCCGAGCGAGT----- 285
Db      749 uArgGlnValHleAlaAspGlnArgAspPro-----LeuGlnGlyProGlyGlnPr 766
Qy      284 -----CGCGCCCGCGCGCG-----CGCGCCCGCGCGCTTGCGCG 250
Db      766 cAlaGlnArgArgGlnAlaArgAlaArgProAlaProValArgGly-GlyThrArgAlaP 786
Qy      249 CAGAGCGCAGAGCCCTCGGGGCG---TCGCCCCCGCGCTCAACGCGGTCACTGAAAAAC 193
Db      786 TGGlyGlnGlyGlyProArgGlyProGluProHleAlaHleAlaThrLeuGlnProA 806
Qy      192 GATCAGAGTAGTGATTTTACCGCGCGCGCGCGAGGCGCG---CGAGCCCCCGCGCG 136
Db      806 rGHis-----LeuHleGlyArgValArgArgAspProGlnA 818
Qy      135 CCCCTCGCGCGGAC---ACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTCT 80
Db      818 lAlaArgGlyAspProArgGlyGluGlyProGlyLeuProHleAlaProValLeuLeu 837

```

RESULT 10

unspecific monooxygenase (EC 1.14.14.1) - common tobacco  
T02995  
N:Alternate names: cytochrome P450 homolog TBP  
C:Species: Nicotiana tabacum (common tobacco)  
C>Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 21-Jul-2000  
C:Accession: T02995  
R:Sugiyura, M.; Sakaki, T.; Yabuuchi, Y.; Ohkawa, H.  
Biochim. Biophys. Acta 1308, 231-240, 1996  
A:Title: Cloning and expression in Escherichia coli and Saccharomyces cerevisiae of a nd  
A:Reference number: Z14816; PMID:96404975; PMID:8809115  
A:Accession: T02995  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-530 <SUG>  
A:Cross-references: EMBL: D64052; NID: G1545804; PIDD: BAA10929.1; PID: G1545805  
A:Experimental source: strain Bright Yellow 2  
C:Genetics:  
A:Gene: CTBP  
C:Keywords: monooxygenase, oxidoreductase

Alignment Scores:  
Pred. No.: 9.39e-06 Length: 530  
Score: 199.00 Matches: 57  
Percent Similarity: 61.1% Conservative: 9  
Best Local Similarity: 52.78% Mismatches: 31  
Query Match: 4.71% Indels: 12  
DB: Gaps: 4

```

US-10-028-952a-4 (1-2333) x T02995 (1-530)
Qy      328 CGCCCAAGTCAAACTCCCACTGACATGTCCTCCGAGCGAGTCGCGCCGCGCGCGC 269
Db      3 ArgProSerGlnThrProHleLeuThrMetSerSerAlaArgyle----- 17
Qy      268 GCGCGCGCGCGCTTGCGCGCGAGAGCGAGCGCGCTTGCGCGCGCGCGCGCGCT--- 213
Db      18 -----GlyLeuGlnAlaSerLeuGlySerLysLysArgGlySerAlaProLeuProIle 35

```

```

Qy      212 CACCGGCTCAGTGAAGAAAAAGATCAGAGTGTGATTTTACCGCGCGCGCGCGCGCG 153
Db      36 HleGlyIleSerLysIleThrLeuLysValValProHleSphArgLeuSerAlaPro 55
Qy      152 GCGAGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 93
Db      56 ThrTrpThrThrProLeuLysSer---PheHleSylValGlyLeuGlnSer---SerSer 73
Qy      92 ACAGGCTCTTCTTCTCCCGCTGATTCGCGCAAGCCCGTTCCTTGCTGTGCTGCTGCT 33
Db      74 ThrGlySerSerPheProAlaAspSerAlaLysProValProLeuValValaSerLeu 93
Qy      32 GATAGTGGTACGAGCAGCGG 11
Db      94 AspSerArg-GlnAspSerGly 100

```

RESULT 11

ALR protein - human  
T03455  
C:Species: Homo sapiens (man)  
C>Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 27-Oct-2003  
C:Accession: T03455  
R:Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano, Oncogene 15, 549-560, 1997  
A:Title: Structure and expression pattern of human ALR, a novel gene with strong homolo  
A:Reference number: Z14954; PMID:97388474; PMID:9247308  
A:Accession: T03455  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-4957 <PRA>  
A:Cross-references: EMBL: AF010404; NID: G2358286; PIDD: AAC51735.1; PID: G2358287  
C:Genetics:  
A:Gene: ALR  
A:Map position: 12  
C:Superfamily: acute lymphoblastic leukemia protein, ALR type  
C:Keywords: alternative splicing

Alignment Scores:  
Pred. No.: 1.75e-05 Length: 4957  
Score: 195.50 Matches: 178  
Percent Similarity: 32.62% Conservative: 65  
Best Local Similarity: 23.89% Mismatches: 271  
Query Match: 4.62% Indels: 232  
DB: Gaps: 33

```

US-10-028-952a-4 (1-2333) x T03455 (1-4957)
Qy      28 CTATCCAGGAAACCAACAGCGGAGCGGCTTGCGGATCAGC----- 75
Db      1546 lIleGlySerProThrThrPro-----AlaGlyLeuSerThrSerAlaAspGlyPheLeu 1563
Qy      76 -----GGAGAAAGAACCCCTGTTAGCTTGAGGCGCGCGCG 111
Db      1564 LysProProAlaGlySerValProGlyProAspSerProGlyGluLeuPheLeuLysLeu 1583
Qy      112 CCGCGCGGTCGTCGCGCGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 171
Db      1584 ProProGln-ValProAlaGlnAlaProSerGlnAspProPheGlyLeuAlaProAla-- 1602
Qy      172 TGAATACCACTACTGTGATCGTTTTCAGTACGCGCGCGCGCGCGCGCGCGCGCG 225
Db      1603 ---TyrProLeu-GluProArgPheProThrAlaProProThrTyProProTyProS 1621
Qy      226 AGCGCGAGGCGCTTCGCTTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 285
Db      1621 erProThrGlyAlaProAlaGlnProProMetLeuGlyAlaSerSerArgProGlyAlaG 1641
Qy      286 CCGCGCTCGCGG-----ACAGTCCAGGTGGGAGGATTTGAAGTGGCGCGGTA 333
Db      1641 LysGln-ProGlyGluPheHleSthrThrProProGlyThrProArgHleGlnProSerThr 1660
Qy      334 CCG-----GTCAACCGTAAACGAGGTGCTTAAAGCGAGCTCAGGAGAGACAGAA 384

```

```

Db      1661 ProMetProPheLeuLys-----ProArgCysProSerLeuAspLeuAlaValPro 1678
      |||  ::::|  |||||  ::|
Qy      385 ACCTCCGTTGAGACAGAGAGGCAAAAGCTGCTTATCTTGAATTTCA----- 432
      |||  |||  |||  |||  |||  |||
Db      1679 GluSerProGlyValGlyGlyValAlaSerClnProLeuSerProProPhe 1698
      |||  |||  |||  |||  |||  |||
Qy      433 -----GTAACAATACAGACCGTGAAGAGGGCTCAACATCTCTCT 474
      |||  |||  |||  |||  |||  |||
Db      1699 GlyGluSerArgLysAlaLeuGluValLysGluLeuGlyAlaSerProSer 1718
      |||  |||  |||  |||  |||  |||
Qy      475 -----GACCTTTGGTTTAAACAGAGAGGTGTACAGAAAATTACCAAGGATTA 525
      |||  |||  |||  |||  |||  |||
Db      1719 TyrClnProProAsnLeuGlyPheValAspSer-----ProSerSergly 1733
      |||  |||  |||  |||  |||  |||
Qy      526 CTGGCTTGTGGCGCCCAAGCGTTATAGCGACGTCGTTTGTATCTTGAATGTCGGCT 585
      |||  |||  |||  |||  |||  |||
Db      1734 ThrHisLeuGlyGlyLeuGluLeuLysThrProAspAlaPheLysAlaProLeuThrPro 1753
      |||  |||  |||  |||  |||  |||
Qy      586 CTTCCTATCATTTGTAAGACAGAAATTCACCAAGCGTTGATTTTCAACCACTAATAGGA 645
      |||  |||  |||  |||  |||  |||
Db      1754 ArgAlaSerGlnValGlnProGlnSerProGlyLeuGlyLeu----- 1767
      |||  |||  |||  |||  |||  |||
Qy      646 ACGTAGAGTGGAGTTAAGCGCTCGTAGACAGAGTTAATTACCTTACTATGATGTGTT 705
      |||  |||  |||  |||  |||  |||
Db      1768 -----ArgPro-GlnGlnPro-----ProProAlaGlnAlaLe 1778
      |||  |||  |||  |||  |||  |||
Qy      706 GTTGCCATGATGTAATCTGCTCAGTACAGAGAGAACCCAGATTACATTTGGTGTATG 765
      |||  |||  |||  |||  |||  |||
Db      1778 ValAspPro----- 1780
      |||  |||  |||  |||  |||  |||
Qy      766 TGTGTGGCTAGAGAGCCAGATGGGCGAAGCTACCATCTGTGGATTAATGACTGAACGCT 825
      |||  |||  |||  |||  |||  |||
Db      1780 ----- 1780
      |||  |||  |||  |||  |||  |||
Qy      826 CTAAAGTCAGATCCCGCCGAGCGGAAAGATAGAGGACAGCGCGCGGAGCCCGTGGTGC 885
      |||  |||  |||  |||  |||  |||
Db      1781 -----SerProProSerHisProAspAlaPheArgProGlySerTyrThrAspPr 1797
      |||  |||  |||  |||  |||  |||
Qy      886 TCGGATAGCCGATCCCGCCGCTGTCCCGCGGCGGCGCGCCGCGCC----- 931
      |||  |||  |||  |||  |||  |||
Db      1797 O---TyrAlaGlnProProLeuThrProArgProGlnProProProProGlnSerCysCy 1816
      |||  |||  |||  |||  |||  |||
Qy      932 -----CTCCAGAGCGCC---CGCGCGCGCGGAGGCGCGCGCCCGCGCGCGCG 981
      |||  |||  |||  |||  |||  |||
Db      1816 AlaLeuProProArgSerLeuProSerAspProPheSerArgValProValSerProGln 1836
      |||  |||  |||  |||  |||  |||
Qy      982 GGAACGGGGGTCGCGAGTGCAGTGCCTTGTCTGTGGAAACGGGCGCGCGCGGAAAGGC 1041
      |||  |||  |||  |||  |||  |||
Db      1836 nSerClnSerSerSerGlnSerProLeuThrPro-----ArgProLeuSerAlaGlnAl 1854
      |||  |||  |||  |||  |||  |||
Qy      1042 GAGCGCGCCCTGCGCCGTCACGACCGACCGACTGTGCG---GAACTCGGCGCTAAACCA 1098
      |||  |||  |||  |||  |||  |||
Db      1854 aPheCysProSerProValThr--ProArgPheGlnSerProAspProTyrSerAspPro 1873
      |||  |||  |||  |||  |||  |||
Qy      1099 CCTTCATCTCCAGT-----CCTCAGCCCTGCGC 1124
      |||  |||  |||  |||  |||  |||
Db      1874 ProSerArgProGlnSerArgAspProPheAlaProLeuHisLysProProArgProGln 1893
      |||  |||  |||  |||  |||  |||
Qy      1125 AAGCTGAGAGAGCCCTTCTCTGACAGAGACGTACACCAACAGAGACCCACTTTGGCAGA 1184
      |||  |||  |||  |||  |||  |||
Db      1894 ProProGluValAlaPheLysAlaGlySerLeuAlaHisThrSerLeuGlyAlaGlyGly 1913
      |||  |||  |||  |||  |||  |||
Qy      1185 GAGCGAGGTGTCGTCATCTCAAGCGGCGGAGGATCTCCCTGAGAGAGCGCGCGCC 1244
      |||  |||  |||  |||  |||  |||
Db      1914 PheProAlaAlaLeuProAlaGlyProAlaGlyGluLysAlaLysValProSer--- 1932
      |||  |||  |||  |||  |||  |||
Qy      1245 AGCACTCTCTCATGTCTGTGTGACGACAA----- 1274
      |||  |||  |||  |||  |||  |||
Db      1933 GlyGlnProProAsnPheValArgSerProGlyThrGlyAlaPheValGlyThrProSer 1952
      |||  |||  |||  |||  |||  |||
Qy      1275 -----GAGAGGCTGTGTATGAG-----GAACCTCCAGAGCAG 1307
      |||  |||  |||  |||  |||  |||

```

```

Db      1953 ProMetArgPheThrPheProGlnAlaValGlyLysProSerLeuLysProProValPro 1972
      |||  |||  |||  |||  |||  |||
Qy      1308 GAGACCTTCTACAGACAGACGCCCACTGTGACAGAGGATGCGGCTTGAGACATTT 1367
      |||  |||  |||  |||  |||  |||
Db      1973 Gln-ProGlyLeuProProProAlaGlyLysAsnSerHisPheGlyProGlyProThrLe 1992
      |||  |||  |||  |||  |||  |||
Qy      1368 G-----ACACCAATTCAGAGGCGCGCGCTC 1394
      |||  |||  |||  |||  |||  |||
Db      1992 uGlyLysProGlnSerThrAsnTyrThrValAlaThrClnLysPheHisProSerGlySe 2012
      |||  |||  |||  |||  |||  |||
Qy      1395 A---GTGGCAAGGCTCT----- 1410
      |||  |||  |||  |||  |||  |||
Db      2012 rProLeuGlyProSerSerglySerThrGlyGluSerTyrClnLeuSerProLeuArgPr 2032
      |||  |||  |||  |||  |||  |||
Qy      1411 -----GTGCCCGGCGCTGTACACTACACAGGACGCG-----AC 1445
      |||  |||  |||  |||  |||  |||
Db      2032 oProSerValLeuProProProAlaProAspGlySerLeuProTyrLeuSerHisGlyAl 2052
      |||  |||  |||  |||  |||  |||
Qy      1446 GACACAGAGATCTCTTGAACCCGAGAACTCATACAGGCGCATGAGAGTGAACGAGA 1505
      |||  |||  |||  |||  |||  |||
Db      2052 aSerGlnArgSerGlyLysThrSerProValGlnLysArg-----Gln 2066
      |||  |||  |||  |||  |||  |||
Qy      1506 GCGTGTGCGCGCTATGAGCGCGGATGACCATTTGTCATGTTCCCTGCAACTACG-- 1563
      |||  |||  |||  |||  |||  |||
Db      2066 uAspProGlyThrGlyMetGly---SerSerLeuAlaThrAlaGluLeuProGlyTyrGln 2085
      |||  |||  |||  |||  |||  |||
Qy      1564 -----TGACCTCATTTAGTGAAGCTGAGGGGACATCTTCCCTCTCAG 1613
      |||  |||  |||  |||  |||  |||
Db      2085 nAspProGlyMetSerGlyLeuSerGlnThrClnLeuGlnLysGlnArgGlnArgGlnArg 2105
      |||  |||  |||  |||  |||  |||
Qy      1614 ACATGGCTTCCTTAT----- 1629
      |||  |||  |||  |||  |||  |||
Db      2105 gLeuArgGlnLeuLeuLlleArgGlnGlnLlleArgAsnThrLeuArgGlnGlnLysGln 2125
      |||  |||  |||  |||  |||  |||
Qy      1630 -----GCTGAAAGAGAGCGCTGGAGCTTAC----- 1656
      |||  |||  |||  |||  |||  |||
Db      2125 uThrAlaAlaAlaAlaAlaGlyAlaValGlyProProGlySerTyrGlyAlaGlnProSe 2145
      |||  |||  |||  |||  |||  |||
Qy      1657 -----ATTGACACTCTTCAGAGAAATGAGACCCCG-----AGTAGAGATGA 1697
      |||  |||  |||  |||  |||  |||
Db      2145 rSerProAlaPheGlnGlnLeuSerArgGlyGlnThrProPheAlaGlyThrGlnAspLys 2165
      |||  |||  |||  |||  |||  |||
Qy      1698 GGCTCTCA-----GGGCTCCCTCCGCGCTTGCGACACTGAGCTTGACCCCAATGACAGC 1751
      |||  |||  |||  |||  |||  |||
Db      2165 sSerSerLeuValGlyLeuProProSerTyrSerLeuSerglyProLleLeuGlyProGlySe 2185
      |||  |||  |||  |||  |||  |||
Qy      1752 AATGGCTGTGATTCACACACATCTTCCTGACATCCCGCGACCCCTCCGACAGACTTG 1810
      |||  |||  |||  |||  |||  |||
Db      2185 r-PheProSerAspAspArgLeuSerArgProProProThrClnAlaThrProSerSerMet 2204
      |||  |||  |||  |||  |||  |||

```

RESULT 12  
 T03454  
 ALR protein - human  
 C/Species: Homo Sapiens (man)  
 C/Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 27-Oct-2003  
 C/Accession: T03454  
 R/Prasad, R.; Zhadanov, A.B.; Sedkov, Y.; Bullrich, F.; Druck, T.; Rallapalli, R.; Yano, Oncogene 15, 549-560, 1997  
 A/Title: Structure and expression pattern of human ALR, a novel gene with strong homolog  
 A/Reference number: Z14954; NID:97388474; PMID:9247208  
 A/Accession: T03454  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Kolecule type: mRNA  
 A/Residues: 1-5262 <PRA>  
 A/Cross-references: EMBL:AF010403; NID:92358284; PID:AACT51734.1; PID:92358285  
 C/Genetics:  
 A/Map position: 12  
 A/Superfamily: acute lymphoblastic leukemia protein, ALR type  
 C/Keywords: alternative splicing

Alignment Scores: 1.75e-05 Length: 5262  
 Pred. No.:

Score: 195.50 Matches: 178  
Percent Similarity: 32.62% Conservative: 65  
Best Local Similarity: 23.89% Mismatches: 271  
Query Match: 4.62% Indels: 232  
DB: 2 Gaps: 33

US-10-028-952a-4 (1-233) x T03454 (1-5262)

QY 28 CTAATCCAGGAAACACAGCCCAAGGAGCGCTTGCGGAATCAGC----- 75  
1851 IIEGlySerProthThPro-----AlaGlyLeuSerThrSerAlaSpGlyPheLeu 1868  
76 -----GGAAAGAAAGACCTGTTGAGCTTGAGCCCGCCGCG 111  
1869 LysProProAlaGlySerValProGlyProAlaSerProGlyGluLeuPheLeuLeu 1888  
QY 112 CCCCCCGGTGTCCTCCCGGAGGCGCCGCGGCGGCGCTGCGGCGCGCG 171  
1889 ProProGln-ValProAlaGlnAlaProSerGlnAspProPheGlyLeuAlaProAla-- 1907  
QY 172 TGAATACCACTACTGATCGTTTTCCTGACTGACCCGCTGAGCGCGGCG--GGCG 225  
1908 ----TyrProLeu-GluProArgPheProThraLProProThrTyrProProTyrPro 1926  
QY 226 AGCCCGGAGGCGCTCTCGCTTCTGGCGCCAGCGCCGCGCGCGCGCGCGCGCGA 285  
1926 eProThrGlyAlaProAlaGlnProProMetLeuGlyAlaSerSerArgProGlyAlaG 1946  
QY 286 CCGCGTCCGCGG-----ACAGTCCAGGCTGGGAGTTGACTGGGCGGTACA 333  
1946 LysGln-ProGlyGluPheHisThrThrProGlyThrProArgHisGlnProSerThr 1965  
QY 334 CCT-----GTCAACGGTAAACGAGCTGCTTAAGCGACGCTCAGAGGAGCAGAA 384  
1966 ProAspProPheLeuLys-----ProArgCysProSerLeuAspAsnLeuAlaValPro 1983  
QY 385 ACCTCCCGTGGAGCAGAGGCGAAAGCTCGCTGATCTTGATTTC----- 432  
1984 GluSerProGlyValGlyGlyGlyLysAlaSerGluProLeuLeuSerProProPhe 2003  
QY 433 -----GTACGATACAGACCGTGAAAGCGGCGCTCAGATCCTTCT 474  
2004 GlyLysSerArgLysAlaLeuGluValLysGluGluLeuGlyAlaSerSerProSer 2023  
QY 475 -----GACCTTTGGGTTTAAAGCAGAGGTCTAGAAAAGTTACACAGGAGTAA 525  
2024 TyrGlyProProAsnLeuGlyPheValAspSer-----ProSerSerGly 2038  
QY 526 CTGGCTTGTGGCGCCAGCGTTCATAGCAGCGCTCTTTTGATCTTCGATCGGCT 585  
2039 ThrHisLeuGlyGlyLeuGluLeuLysThrProAspValPheLysAlaProLeuThrPro 2058  
QY 586 CTTCCTATCTTTGTAAGCAGAAATTACCAAGCTTGATTTGTTACCCAGTAATAGGA 645  
2059 ArgAlaSerGlnValGluProGlnSerProGlyLeuGlyLeu----- 2072  
QY 646 ACGTAGCTGAGATTAGACCGTCGTGAGACAGGTTAGTTTACCTACTGATGATGTT 705  
2073 -----ArgPro-GlnGluPro-----ProProAlaGlnAlaLe 2083  
QY 706 GTTGCCATGTGATCTGCTCACTAGTACGAGAGAACCGCAGGTTCAACATTTGGTGTAG 765  
2083 ValAlaPro----- 2085  
QY 766 TGCTTGCTAGAGCAATGGGCGAAGTACATCTGGGATTATGACTGAAGCCT 825  
2085 ----- 2085  
QY 826 CTAAGTCAGAAATCCCGCCAGCGGAAAGATACCGCAGCGCGGAGGCTCGGTTGGCC 885  
2086 -----SerProProSerHisProAspIlePheArgProGlySerTyrThrAspPr 2102  
QY 886 TCGGATAGCGGCTCCCGCGCTGCCCCGCGGCGCGCCGCC----- 931

DB 2102 o--TyrAlaGlnProProLeuThrProArgProGlnProProProGluSerCysCy 2121  
QY 932 -----CCTCCAGCGGCC---CCGCGCGCGGAGGCGGCGCGCGCGCGCGCG 981  
2121 SalAlaLeuProProArgSerLeuLeuProSerAspProPheSerArgValProValSerPro 2141  
QY 982 GAGCCGAGGCTCCGAGTGCAGTGCCTTCTGCTGAGAAACCGGAGCGCGCGGAGAAAGC 1041  
2141 nSerGlnSerSerSerGlnSerProLeuThrPro-----ArgProLeuSerAlaGlu 2159  
QY 1042 GCGCGCGCGCGCGCGCGCTCAGACGACCGGACGTTGCGG3---GAACCTGAGCTAAACCA 1098  
2159 AsnProProProSerProValThr--ProArgPheGlnSerProAspProTyrSerArgPro 2178  
QY 1099 CCTTCATCTTCAGT-----CCTCAGCTTGGC 1124  
2179 ProSerArgProGlnSerArgAspProPheAlaProLeuHisLysProProArgProGln 2198  
QY 1125 AAGTAGAGAGCCCTTCCTGACAGAGAGCTCACCCAGAACCAAGACCACTTGGCAGA 1184  
2199 ProProGluValAlaPheLysAlaGlySerLeuAlaHisThrSerLeuGlyAlaGly 2218  
QY 1185 GAGCCAGCTGCTCCATTCAGAGCCCGAGGCGAGATCTCCCTGAGAGAGCGCGCGCC 1244  
2219 PheProAlaAlaLeuProAlaGlyProAlaGlyGluLeuHisAlaLysValProSer-- 2237  
QY 1245 AGCACTCTTCAGTCTGTCGTCAGGAGCAAA----- 1274  
2238 GlyGlnProProAsnPheValArgSerProGlyThrGlyAlaPheValGlyThrProSer 2257  
QY 1275 -----GAGAGGCTGTGTATAG-----GAACCTCCAGACAG 1307  
2258 ProMetArgPheThrPheProGlnAlaValGlyGluProSerLeuLysProProValPro 2277  
QY 1308 GAGACCTTCTTACAGAGAGCCCGGCTGTCAGACAGCAAGTCTGCTGAGCAGAT 1367  
2278 Gln-ProGlyLeuProProProHisGlyLysLeuSerHisPheGlyProGlyProThrLe 2297  
QY 1368 G-----ACCACCACTTGAAGGCGGCGCTC 1394  
2297 uGlyLysProGlnSerThrAsnTyrThrValAlaThrGlyAsnPheHisProSerGlySe 2317  
QY 1395 A--GTGGCGAGGCTCT----- 1410  
2317 rProLeuGlyProSerSerGlySerThrGlyGluSerTyrGlyLeuSerProLeuArgPr 2337  
QY 1411 -----GTGCGCGTGCCTGTACGACTACAGACAGCAGCG-----AC 1445  
2337 oProSerValLeuProProProAlaProAlaProGlySerLeuProTyrLeuSerHisGlyAl 2357  
QY 1446 GACACAGAGATCTCTTGAAGCCCGAGAACCTCATCAAGGAGCTGAGGTGATGAGAA 1505  
2357 AsnGlnArgSerGlyLeuThrSerProValGluLysArg-----G 2371  
QY 1506 GCGCTGTCGCGCTGAGCTTGGCGGAGGATTTGGCATGTTCCCTGCCAATACG-- 1563  
2371 uAspProGlyThrGlyMetGly---SerSerLeuAlaThrAlaGluLeuProGlyThrG 2390  
QY 1564 -----TGAGCTCATTAAGTGAAGCTGAGGAGCATCTTGCCCTTCCCTTCAG 1613  
2390 nAspProGlyMetSerGlyLeuSerGlnThrGluLeuGluLysGlnArgGlnArgGlnAr 2410  
QY 1614 ACATGCTTCTCTTNT----- 1629  
2410 GLeuArgGluLeuLeuIleArgGlnGlnIleGlnArgAsnThrLeuArgGlnGlnLysG 2430  
QY 1630 -----GCTGAAGAGAGAGGCTTGAAGTAC----- 1656  
2430 uThrAlaAlaAlaAlaAlaGlyAlaValGlyProProGlySerTrpGlyAlaGluProSe 2450  
QY 1657 -----ATTGACACTCTTCAGAGATAGAGACCCCG-----AGTGAAGATGA 1697

Db 2450 SerProAlahegIugInIeuSerArgGlyGlnThrProheAlaagIthrgInaaply 2470

Qy 1698 GGCTCA-----GGCTCCCTCGCGCTGGCAACACAGCTTCAACCCCAATGCAGC 1751

Db 2470 .eSeSerIeuValaIeuProPheSerLysIeuSerLysProIleuGlyProGlySe 2490

Qy 1752 AATGCGCTGGTGATCCACACATCTTCTTCGATCCCGCAGCCCTCCGACAGCTTG 1810

Dh 2490 r-PheProSerAspAspArgIeuSerIaArgProProProAlaThrProSerSeret 2509

RESULT 13  
 TC4163  
 DNA-binding protein SES - rat  
 C/Species: Rattus norvegicus (Norway rat)  
 C/Date: 27-Aug-1995 #sequence\_revision 19-Oct-1995 #text\_change 20-Jun-2000  
 C/Accession: J04163; PC4040  
 C/Sunuki, E.; Kojima, N.; Yoshimura, K.; Uemura, K.; Obata, K.; Akagawa, K.  
 J. Biochem. 118, 122-128, 1995  
 A/Title: Cloning and sequence analysis of cDNA for a possible DNA-binding protein SES in  
 J/Reference number: J04163; MUID:96015159; PMID:8537300

A:/Accession: J04163  
A:/Molecule type: mRNA  
A:/Residues: 1-825 <SUB>  
A:/Cross-references: DDBJ:J037934, NID:G531260, PIDN:BAA07153.1, PID:G531261  
A:/Experimental source: Brain  
A:/Accession: PC040  
A:/Molecule type: Protein  
A:/Residues: 230-455 <SUB>

C:Comment: This protein has an abundance of arginine, a glycine-rich region and a proline-rich region.  
C:Superfamily: collagen alpha 1(I) chain; fibrillar collagen carboxyl-terminal homology  
C:Keywords: nerve; phosphoprotein  
E:435-443/Region: nuclear location signal  
E:725-731/Region: proline cluster

Alignment Scores:

Pred. No.:	2.2e-05	Length:	82
Score:	193.50	Matches:	19
Percent Similarity:	30.39%	Conservative:	70
Best Local Similarity:	22.36%	Mismatches:	28
Query Match:	4.58%	Indels:	3
DB:	2	Gaps:	43

US-10-028-952A-4 (1-2333) X JC4163 (1-825)

QY	2261	GGCACACACTGGGGCTGCAGACGATCGAGTGA-----CCGCGTGAGAGAACTGT	2211
Db	113	GlyThrGlyTTPGilyLeuSerGilyLeuSerProThrAlaLeuSerArgAspGilyTyrglu	132
QY	2210	TGTGTCTGCCGCTTTTGAATACCAAGGTGGAGCTTGCCATCTGCATCCCACTT--	2154
Db	133	ThrGluThrProPhe-----SerProGluGlyAlaPheProGlyGilyGlyProAlaGlu	150
QY	2153	-----CCCATAGCCAGGACAGGACAGAGAAATGSGTGT-----	2118
Db	151	GluGluGlyValProArgProArgAlaProProGluProPheProAlaProArg	170
QY	2117	-----GGAGACAGACAGACAGGCTCCACACAGA	2091
Db	171	ProProPheAspProGlyProLeuProLeuProGlySerGingluys--ProThrPhe	189
QY	2090	CAAAATTCCTCTGCTGCCAAACCAACCATGATCCACTCTGACTTGGTACAAACTCTGCTAA	2031
Db	190	ValValGlnValSerThrGluGlnLeuLeuMetSerThrGilyGly-----Cysasp	206
QY	2030	AAACAAATCTCTMACGTTCACTGTTCCCAAGGT-----CAT	1995
Db	207	LysGlu-----ProProArgGilyGlnGlyValAspThrArgGilyAspArg	221
QY	1994	TCTAAACAGTGTGTGCCACCAACCCCTTAAAGACAAAGAAAAAAGCAAAATAA	1935
Db	222	ThrGinglu-GilyGilyGluysProArgGluGlnArgGluGlyProArgProGinglu	241

QY	153	TCGAGAAAGGACGACGAGCCTCCCAATGCGCCCTGCTCAGAGACCTTCCCGCAGCTTGG	1875
Db	241	YPR0A8B11EProGlyInGlnInGlnInGlnUserProGlnInGlnInUProSerSerGlyUArgG1	261
QY	1874	CAGT-----GSCCACTCAGGAGCTTGGCCACAGACA	1845
Db	261	YASpSerValGlyUArgGlnUArgSerProGlyYH1GslUg1Y-----G1	277
QY	1844	GCGGTTGGCTCAGATGCT-----GTCAGGGGACAGAGCCAGCTGTCG3--	1799
Db	277	UG1GlyGlyGlyUtrProGlyIleSerGlyGlyUArgArgGlnUserProGlyUtrPgl	297
QY	1798	-----GAGGATCGGGGATGACAGAAAGATGTGTGGAAATCACAGGCCATT	1752
Db	297	YAlAspValProArgGlyUArgGlyG1Ug1YAlAg1YUtrPglYSerAspValPro--	316
QY	1751	GCTGCATTGGGGTGAAGAGCTGATGCTGGCAAGCGGAGGGAGCCTGAGGCTCATCC	1692
Db	317	-----LYAspArgGlyG1Ug1Yg1YArgUtrPglYProGlnUAlaGlnG1	333
QY	1691	TCAC-----TGGGGGCTTATTCCTGGAAGAGTGTGATGCTCACT	1650
Db	333	UH1GslYGlnUAlaArgAspTrpThrSerGlnUser-PrArgThrLeuGlyUAspA	353
QY	1649	CCGAGGCTCCTCTTCAGACAAATAGGAACCATGTCTGAGAGGGGAAGGCAAGATGTG	1590
Db	353	IAArgAspTrpGlyYSerSerSerArgAspAla-----AlaGlySerSerProCysA	370
QY	1589	CCCTC-----AGCCTCATCA-----	1573
Db	370	IAlaUArgGlySerLeuUAlaProGlnArgLeuGlyAspGlyProTrpProAlaTrpProS	390
QY	1572	-----TANAGTCCACGTATGTTGGACG	1551
Db	390	eRProGlnUArgGlnProGlyYProArgAspArgValGlnUserProArgGlnUtrPglYg	410
QY	1550	GAACA-----TCCCAAAATGGCATTCGGCCATA	1521
Db	410	LYThrGlnUserProArgGlyYTrpGlnUAlaGlyProArgGlnUtrPglYProSerProGlyg	430
QY	1520	GCCAGCCACAGCAGCTTGTGTGATCACCTGTGAGGCCCGTGAATGAGTTCTCGGGGTCAA	1461
Db	430	LYArgGlyAspGlyYProArgArgArgProArg-----LYAspArgGlyYArgL	446
QY	1460	GGAATATCTGTGTGTCGCGCTGCTGTCGTATGCTGACAGGGACGCGACAGAGCCCTTG	1401
Db	446	Ysg1YArgMetGlyYArgGlnLeuUtrThrAlaTrSer-----AlaSerA	462
QY	1400	CCCATAGAGCCCTGCGCCCTGATGTGTGTGTCATATGCTCAGAGCCACAGACTTGTCTG	1341
Db	462	IAThrGlyGlyYProAlaGlnUAlaGlyAlaSerAlaProGlnUg1Yg1n--AlaGlyg	481
QY	1340	CTGACACAGTGGGGGCTGCTGTGTAAGAGTCTCTGCTGAGAGTTCTCATACACAGC	1281
Db	481	LYg1YProArgGlyYArgAlaArg-----	488
QY	1280	CTCTCTTTCTGCTGCACCAACATGAGAGAGTCTGGGCGCGGCTCTCAGACAGGAG	1221
Db	489	-----G1YProArgGlnGlnUAlaArgArgYH1GslYProGln-Arg--	502
QY	1220	ATCTGCGCTGGGCTTGATGACGACAGAGTGGCTCTCTGCAAAAGTGGGCTCTGCTTG	1161
Db	502	-----	502
QY	1160	GGTGAGCTGCTTTCGACAGAGGGGCTCTCAGCTTCCAGGCTGAGACTGAGAT--	1104
Db	503	-----ArgArgGlyYProProGlnUAlaGlyGlnUg1Yg1YProGlyYAspAla	517
QY	1103	-----GGAGTGTGTTTAGGCGCAGGTTCCCAAGAACGTGCGGGTGCATGACGG	1066
Db	518	ThrLeuValLeuGlyLeuGlyYThrThrSerGlyGlnUArg-----AlaAspGln	534

```

Db      535  ....
        536  SerGlnThrLeuProAlaLeuAlaGlyAlaProThrAla-----HisAlaHisAlaVal 552
QY      995  CCGGACCCCGGTCGCGCGCGCGG-----CGGGGACCGCGCCCTCCGCGCGC 948
        553  ProGlyProGlyProAlaAlaAlaThrLeuGlyGlyArgGly-Arg----- 567
QY      947  GGGGGCGGCGTGAAGGGGGGGCGCGCGCGCGCGG-----ACAGGCGGGGAC 897
        568  -ArgGlySerThrPaGlyGlyArgArgGlyGly-GlyAlaAlaValAsnGlyGlyGly 587
QY      896  CGGCTATCCGAGCCACCGAGG-----CTCCGGCGCGCTCCGCTATGTCCTCCGCTGG 843
        587  TgGlyGlyArgGlyArgGlyArgGlyGlyArgArgGlySerGlyLeuSer----- 603
QY      842  GCGGGATTCTGACTTAGAGCGCTTCATGATATCCACAGATGATGATGCGCCCAT 783
        604  -----G 604
QY      782  GGCTCCTCAGCCAGACATACACCAATGTCGACCTGCTCTCTCTCTCTCTCTCTCT 723
        604  LThrArgGlnAspAlaGlySerProSer-----AlaArgArgGlyGlnG 619
QY      722  AGGATTACCATGCGCAACACATCATCATGATAGGTAATACTGCTGCTCAGACGGT 663
        619  LArgArgArgGlyHisGlyPro-----ProAlaAlaGlyAlaAla 632
QY      662  CTATCCACGCTCAGCTTCCTTATGATGATGATGATGATGATGATGATGATGATG 603
        632  LAsnValSerThrArgGly----- 638
QY      602  TTCACATGATGAGAGCCGACATCAGATCAAGATCAAAAAGCA----- 559
        639  -----ArgArgAlaArgGlyGlnArgThrGlyGlnGlyAlaGlnAspGlyLeuL 555
QY      558  -CGTGGCTATGACGCTTGCGCGCGCACCAAGCATATCCCTGCTGTAATTTTCT 504
        655  EurProArgGlyArgAspArgGlyLeuProLeuArgProGlyAspSer----- 669
QY      503  GACACCTCTCTTAAACCAAGAGTCAGAGATCCTGAGCGCCCGC-----TTTCAGC 447
        670  -----AsnGlnArgValGlnArgProGlyHisProArgGlyGlyHisG 684
QY      446  GTCTGATTCG-----T 435
        684  LAlaAlaLeuAlaProSerAlaProAspAlaSerProProHisProArgArgTyrP 704
QY      434  ACTGAAATCAAGATCAAGCGAGCTTTGCCCTTCTCTCCACGGAGGTTTCTGCTCT 375
        704  AlSerGlnGlnArgGlnArgLeuTyrArgGlnPheArgValGlyGlyGlyPheProP 724
QY      374  CTGAGAGTCCCTTAGGACACCTGGCTTACCGCTTACAGGTACCGCCAGTCAAC 315
        724  T-----ProProThrArgPro-----ProProValLeuL 734
QY      314  TCCCACTGCGACTGTCCTCCGAGCGGTCGCGCGCGCGCGCGCGCGCGCGCGCT 255
        734  EurProLeuLeuArgGlyLeuThrCysAlaGlyAsp---ProGlyAlaSerArgProGlySerA 753
QY      254  GGGGCCGAAGCGAGAGCCCTCGGGGCTGCGCCCGCTCAACGGGT----- 205
        753  TgArgProAlaArgArgProArgGlyGlyLeuLeuThrProGlnArgProSerProPheAlaP 773
QY      204  -----CACTGAAAA 195
        773  roGlnGlnGlnGlyLeuArgAlaGlySerCysValAspArgGlyAlaAlaLeuProAspT 793
QY      194  ACAGTCAAGTAGTGTATTTCACCGCGCGCGCGC-----A 159
        793  hrAspThrAlaSerGlyGlyValProGlnAlaGlyProSerLeuSerSerThrMetCysG 813
QY      158  GGGCGCGCGAGACCCCGCGCGCGCGC 133

```

```

Db      813  InMetGlyArgProArgProSerPro 821
RESULT 14
Q0BE3
BHLF1 protein - human herpesvirus 4 (strain B95-8)
C/Species: human herpesvirus 4, Epstein-Barr virus
C/Date: 25-Feb-1985 #sequence_rev150n 25-Feb-1985 #text_change 23-Aug-1997
C/Accession: A03742
R/Author: A.T.; Deininger, P.L.; Farrell, P.J.; Barré, B.G.
Mol. Biol. Med. 1, 21-45, 1983
A/Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus
A/Reference number: A93065; MUID:85055713; PMID:6092825
A/Accession: A03742
A/Molecule type: DNA
A/Residues: 1-660 <BAM>
R/Baser, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.;
Nature 310, 207-211, 1984
A/Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
A/Reference number: A03794; MUID:84270667; PMID:6087149
A/Contents: annotation; protein coding region
C/Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399-5;
C/Superfamily: human herpesvirus 4 BHLF1 protein

Alignment Scores:
Pred. No.: 3.05e-05 Length: 660
Score: 191.50 Matches: 169
Percent Similarity: 27.86% Conservative: 38
Best Local Similarity: 22.75% Mismatches: 213
Query Match: 4.53% Indels: 324
DB: 1 Gaps: 34

US-10-028-952a-4 (1-2333) x Q0BE3 (1-660)
QY      24  CTTACTTTCAGCGCAACCAAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 83
        95  ProAlaGlnGlnAlaAspHisAlaHisSerAsnProThrGlyCysSerAspProGln 114
QY      84  AGACCTGTGAGGTTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 131
        115  Arg-----SerProArgThrArgGlnAla 122
QY      132  -----GGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 158
        123  GlyTyrAlaLeuGlyGlnGlySerAlaGlyLeuGlySerArgGly-----ProArgPro 140
QY      159  TCGCGCGCGCGCGGAGAAATCACTACTGATCGATTTTCTCACTGACCGG----- 209
        141  --HisProAlaPheGlnValGlnTyrSerAlaArgAsnProGlyCysProArgThrTyr 159
QY      210  GTGAGCGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 260
        160  ArgArgArgSerGlyAlaGlnArgGlyHisProProProGlyAlaGlyGlnArgProSer 179
QY      261  ---CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 314
        180  GlyProThrGlyGlyArgProAlaAla-ProGlyAlaProGlyThr---ProAlaAlaP 198
QY      315  GTTGAAGTGGGCGGTACACTGTCACAAACGGTAACGAGGTGTCCTAAGCGAGCTCAG 374
        198  cGlyProGlyGlyGlyAlaAlaVal-----ProSerG 209
QY      375  GAGGACA---GAACCTCCGCTGAGAGAGAGGCAAAAGCTGCTGATCTTATTTTC 431
        209  yAlaThrProHisProGlnArgGly----- 217
QY      432  AGTACGATACAGACCGTGAAGAGCGGCGCTCAGATCTTCTGACCTTTGGTTTAA 491
        218  -----SerGlyProAlaAspProProAlaAlaAlaArgLeuP 230
QY      492  GCAGAGAGTGTCAAA---AAGTTACCAAGAGGATACGTGCTTGTGGCGCGCAAGGTT 548
        230  oProGlnArgGlnGlnProArgProGln----- 240

```

QY	549	CATAGCGACGTCGGTTTTTGA	TCTTGATCGTGATGTGGGCTCTTCTGTA	TCATGTGGAAGCA	608
Db	241	-----	-----	-----AspLeuAlaAlaGlnAr	247
QY	609	TTCAACCAAGCGTGGATGTGTT	CAACCCATAAGAGAACTGAAGTGGATAGACCGTC	668	
Db	247	GCYSPPro-----	-----	-----AlaGlyProProPoth	255
QY	669	GTGAGACAGGTAGTTTACCTT	ACTATGATGCTGTGTTGGCATGGAACTCCGCTCAG	728	
Db	255	T-----	-----	-----	255
QY	729	TACAGAGAAACCGACGATTG	CAGCATTTGGTGTAATGTCTTGAGTGAAGACCAATGGG	788	
Db	255	-----	-----	-----	255
QY	789	GCGAAGCTACCATCTGTGGGA	TTTATGACTGAACCGCTTAAGTCAAAATCCCGCCAGGC	848	
Db	255	-----	-----	-----	255
QY	849	GGAACGATACGACGACCGCGAG	-----	-----CTCGAGTTGGGCTGGA	890
Db	256	---ArgSerGlyAlaAlaAlaGln	ArgThnHisArgArgProProGlyCysProArgSe	274	
QY	891	TAGCCGCTCCCGCCGCTGTCCCG	-----	-----	915
Db	274	ALAArgAsnProGlyCysPro	ArgThrTrpArgArgArgSerGlyAlaGlnArgGly	294	
QY	916	-----	-----CGGCGGCGCGCC	CCCCCTCCACAGCGGCCCGCGCGCGC	953
Db	294	SProProProGlyAlaGlyGln	ArgProSerGlyProThrGlyGlyArgProAlaAlaAr	314	
QY	954	GAGAGGCGCGTCGCCCGCCCG	CGCGCGAGACCGGAGTCCCGTGCAGATCCCTTCGTC	1013	
Db	314	OGlyAlaProGlyThrProAla	AlaProGlyProGlyGlyAlaAlaValProSerG	334	
QY	1014	CTGG-----	-----GAAACGGGCGCGCGCG	GAAGGCGCGCCCGCTCCGCGCGTCA	1061
Db	334	YAlaThrProHisProGln	ArgGlySerGly-----ProAlaAspProAlaAlaAlaAr	353	
QY	1062	CGACCGGACGCTTGCGGGAAC	CT-----	-----	1086
Db	353	GLeuProProGlnArgGln	GlnProArgLeuProGlnAspLeuAlaAlaGlnArgC	373	
QY	1087	GCGCGCTAAACCACTTCAT	CTCCAGTCCCTCAGCC-----	-----TGCAAGCTG	1130
Db	373	SProAlaGlyProProPoth	ArgSerGlyAlaAlaAlaGlnArgThnHisArgArP	393	
QY	1131	AGGAGCCCGCTTCCTGCA	AAAGCAAGCTACCAACAGAGACCCACTTTGGAGAGAGGCA	1190	
Db	393	roProGlyCysProArgSer	AlaArgAsnProGlyCysProArgThrTrpArgArgArS	413	
QY	1191	GCTGCTGCGCATCTCAAG	CGCCAGGCGCAGATCTCCCTGCTGAAGACCGCGCCAGCACT	1250	
Db	413	er-----	-----GlyAlaGlnArgGlyHis	-----SProProProGlyAlaGlyGlnArgProS	429
QY	1251	CCTCCATGCTCGTGCA-----	-----	-----G	1268
Db	429	erGlyProThrGlyGlyArg	ProAlaAlaArProGlyAlaArProGlyThnProAlaArProG	449	
QY	1269	GCAAGAGAGAGGCGTG	TATAGAGAACCTCCAGAGCAGAGACCTTCTACAGAGAGCC	1328	
Db	449	lyProGlyGlyGlyAla-----	-----	-----AlaValP	457
QY	1329	CCACTGCTGCGACGAC	CAAGAGTGCTGGCTCTAGACACATTTGACCAACCAAT-----	1378	
Db	457	roSerGlyAlaThr-----	-----	-----ProHisProGlnArgG	467
QY	1379	--TAGAGGCCAGGAGCT	CAGTGGCGCAAGGCTCTGTGCGCG-----	1417	
Db	467	lySerGlyProAlaAsp	ProProAlaAlaAlaArgLeuProProGlnArgGlnGlnProA	487	
QY	1418	-----	-----TGCCCTGTA-----	1426	

```

Db      487  rglEupProGlnAspLeuAlaAlaAlaGlnArgCysProAlaGlyProProProThrArgS 507
QY      1427  -----CGACTACCAAGCGAG-CCGACGACACAGAGAT 145
Db      507  erG1yAlaAlaAlaGlnArgThrHisArgArgProProGlyCysProArgSerAlaArgA 527
QY      1457  CTCCTTTGACCCCGGAAGACCTCATCAGCGGCATCG----- 149
Db      527  smProGlyCysProArgThrTrpArgArgArgSerGlyAlaGlnArgGlyHisProProP 547
QY      1492  --AGGTGATCGACGAGAGCGCTGAGCGCTGCGCTGATGGCCGAGTGGCCATTGGCATGTT 154
Db      547  rGd1yAlaGlyGlnArgProSerGlyProThrGlyGlyArgProAla-----AlaProG 565
QY      1550  CCTGCGCAACTCAGCGAGGAGCTCATATTGAGTGAAGCGTGAAGGCACATCTGCGCTCCCTC 160
Db      565  lYAlaProGlyThrProAlaAlaProGlyProGlyGlyAlaAlaAlaValProSerGlyA 585
QY      1610  TCAGACATGCTGCTTCCTTATTTGCTGGAAGAGGAGCGCTGGAGTGCATTCAGCACTCTT 166
Db      585  la-----ThrProHisProG 590
QY      1670  CCAGGATATGACCCCGCAGTGAAGATGAGGCTTAGG---CTTCCTCCGCGTTGGCAAA 172
Db      590  lnrArgGlySerGlyProAlaAspProProAlaAlaAlaArgLeuProProGlnArgGlnG 610
QY      1727  CTCAGCGCTGTACCCCAATGACGAGCATGAGCGCTGTGATTCACACATCTCTCTGCAT 178
Db      610  lnrProArgLeuProGlnAspLeuAlaAlaAla-----GlnArgCysProAlaG 626
QY      1787  CCCCAGACCTTCCCAAGACGCTTGCTTCGCTTCGCCCTGACAGATACTGAGCCAAAGCCCTG 184
Db      626  lYPro---ProProThrArgSerGlyAlaAlaAlaGlnArgThrHisArgArgProProG 645
QY      1847  CCTGT 1851
Db      645  lYcys 646

RESULT 15
EDBB11
Immediate-early protein IE110 - human herpesvirus 1 (strain 17)
C:Species: human herpesvirus 1
C:Date: 31-Mar-1988 #sequence_revision 31-Mar-1988 #text_change 17-Mar-2000
C:Accession: A29152
R:Perety, L.J.; Rixon, F.J.; Everett, R.D.; Frame, M.C.; McGeoch, D.J.
J. Gen. Virol. 67, 2365-2380, 1986
A:Title: Characterization of the IE110 gene of herpes simplex virus type 1.
A:Reference number: A29152; MUID:87059760; PMID:3023529
A:Molecule type: DNA
A:Residues: 1-775 <EB>
A:Cross-references: GB:X04614; NID:G59832; PIDD:CAA28285.1; PID:G55833
C:Genetics:
C:Insertions: 19/3; 242/1
C:Superfamily: herpesvirus immediate-early protein IE110; RING finger homology
C:Keywords: DNA binding; early protein; transcription regulation; zinc finger
F:112-162/Domain: RING finger homology <RNG>
F:116-156/Region: zinc finger CHC4 motif

Alignment Scores:
Pred. No.: 3.32e-05 Length: 775
Score: 191.00 Matches: 148
Percent Similarity: 31.07% Conservative: 49
Best Local Similarity: 23.34% Mismatches: 192
Query Match: 4.52% Indels: 246
DB: 1 Gaps: 30

US-10-028-952A-4 (1-2333) x EDBB11 (1-775)
QY      85  GACCGTTGAGGTTGGGCCCCCGGCGGCGGCTGTCGCCGCGAGAGGCCCGGGGCG 144
Db      240  AspAspAlaAspTyrValProProAlaArgArgArgArgArgArgAlaProProAlaArgS 259

```







```

1 FEATURE:
2 NAME/KEY: UNSURE
3 LOCATION: (42)
4 OTHER INFORMATION: Xaa is any amino acid or may be absent
5 FEATURE:
6 NAME/KEY: UNSURE
7 LOCATION: (45)
8 OTHER INFORMATION: Xaa is any amino acid or may be absent
9 FEATURE:
10 NAME/KEY: UNSURE
11 LOCATION: (49)
12 OTHER INFORMATION: Xaa is any amino acid or may be absent
13 FEATURE:
14 NAME/KEY: UNSURE
15 LOCATION: (65)..(66)
16 OTHER INFORMATION: Xaa is any amino acid or may be absent
17 FEATURE:
18 NAME/KEY: UNSURE
19 LOCATION: (82)
20 OTHER INFORMATION: Xaa is any amino acid or may be absent
21 FEATURE:
22 NAME/KEY: UNSURE
23 LOCATION: (99)
24 OTHER INFORMATION: Xaa is any amino acid or may be absent
25 FEATURE:
26 NAME/KEY: UNSURE
27 LOCATION: (115)..(116)
28 OTHER INFORMATION: Xaa is any amino acid or may be absent
29 FEATURE:
30 NAME/KEY: UNSURE
31 LOCATION: (122)
32 OTHER INFORMATION: Xaa is any amino acid or may be absent
33 FEATURE:
34 NAME/KEY: UNSURE
35 LOCATION: (132)
36 OTHER INFORMATION: Xaa is any amino acid or may be absent
37 FEATURE:
38 NAME/KEY: UNSURE
39 LOCATION: (136)
40 OTHER INFORMATION: Xaa is any amino acid or may be absent
41 FEATURE:
42 NAME/KEY: UNSURE
43 LOCATION: (149)
44 OTHER INFORMATION: Xaa is any amino acid or may be absent
45 FEATURE:
46 NAME/KEY: UNSURE
47 LOCATION: (165)..(166)
48 OTHER INFORMATION: Xaa is any amino acid or may be absent
49 FEATURE:
50 NAME/KEY: UNSURE
51 LOCATION: (174)
52 OTHER INFORMATION: Xaa is any amino acid or may be absent
53 FEATURE:
54 NAME/KEY: UNSURE
55 LOCATION: (179)
56 OTHER INFORMATION: Xaa is any amino acid or may be absent
57 FEATURE:
58 NAME/KEY: UNSURE
59 LOCATION: (182)
60 OTHER INFORMATION: Xaa is any amino acid or may be absent
61 FEATURE:
62 NAME/KEY: UNSURE
63 LOCATION: (199)
64 OTHER INFORMATION: Xaa is any amino acid or may be absent
65 FEATURE:
66 NAME/KEY: UNSURE
67 LOCATION: (205)
68 OTHER INFORMATION: Xaa is any amino acid or may be absent
69 FEATURE:
70 NAME/KEY: UNSURE
71 LOCATION: (215)..(216)
72 OTHER INFORMATION: Xaa is any amino acid or may be absent
73 FEATURE:

```

NAME/KEY: UNSURE	LOCATION: (232)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
FEATURE:		
NAME/KEY: UNSURE	LOCATION: (236)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
FEATURE:		
NAME/KEY: UNSURE	LOCATION: (249)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
FEATURE:		
NAME/KEY: UNSURE	LOCATION: (265)..(266)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
FEATURE:		
NAME/KEY: UNSURE	LOCATION: (274)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
FEATURE:		
NAME/KEY: UNSURE	LOCATION: (282)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
FEATURE:		
NAME/KEY: UNSURE	LOCATION: (287)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
FEATURE:		
NAME/KEY: UNSURE	LOCATION: (299)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
FEATURE:		
NAME/KEY: UNSURE	LOCATION: (313)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
FEATURE:		
NAME/KEY: UNSURE	LOCATION: (315)..(316)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
FEATURE:		
NAME/KEY: UNSURE	LOCATION: (332)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
FEATURE:		
NAME/KEY: UNSURE	LOCATION: (349)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
FEATURE:		
NAME/KEY: UNSURE	LOCATION: (365)..(366)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
FEATURE:		
NAME/KEY: UNSURE	LOCATION: (380)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
FEATURE:		
NAME/KEY: UNSURE	LOCATION: (382)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
FEATURE:		
NAME/KEY: UNSURE	LOCATION: (399)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
FEATURE:		
NAME/KEY: UNSURE	LOCATION: (415)..(416)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
FEATURE:		
NAME/KEY: UNSURE	LOCATION: (432)	Xaa is any amino acid or may be absent
OTHER INFORMATION:		
FEATURE:		
NAME/KEY: UNSURE		

```

LOCATION: (449)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (465)..(466)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (482)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (499)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (515)..(516)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (532)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (542)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (544)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE

```

## Alignment Scores:

```

Pred. No.: 1.76e-220 Length: 793
Score: 3664.00 Matches: 696
Percent Similarity: 87.77% Conservative: 0
Best Local Similarity: 87.77% Mismatches: 80
Query Match: 86.62% Indels: 17
Gaps: 14

```

US-10-028-952a-4 (1-233) x US-10-028-952a-9 (1-793)

```

QY 2 CAGGAGATCCCATGTCCTACTATCCAGCAACCAACGCAAGGAGGCT 61
Db 1 HistuileProtiValProthrtYrProAlaYsProGln*****GluArgAla 20
QY 62 TGGCGAATCAGCGGGAAGAGACCTGTAGC----- 97
Db 21 TrpArgangInaYgGlyLysLethrLeuSerHrileuValTrpHisGlyGlu 40
QY 98 -----TTGGGCCCCCGGGGCCCCCGGCTTCGCCGCGAG 133
Db 41 Thr**GluVal**AsnLysTrp**AlaProGlyAlaProProValSerProArg 60
QY 134 GGGCGGGGGCGGGGCGCGGCGCGGCGCGGCGGTAATAACACTACTGATCG 193
Db 60 YLalArgLysLys*****ArgProGlyGlyProProValLysYrHisLysSer 80
QY 194 TTTTTCACCTGACCCGGTGAAGCGGGGGGCGAGCCCGGAGGGCTTCGCTTGC 253
Db 80 Gphe**ThraPProValArgArgGlyGluProArgGlyAlaLeuAlaSer**Al 100
QY 254 CAAGGCGCGGCGCGGCGCGGCGCGGCGCGGCGGCGGCGGCGGCGGCGGCGG 313
Db 100 alyeArgProAlaAlaArgArgProGlyAlaLthrHisSerLys*****AlaArg 120
QY 314 AGTTGACTGGGCGCGGTAACACTGTCAACCGTAACGAGGTGTCTTAAGCGAGCT 373
Db 120 YVal**LeuGlyArgLysLysGlnThrVal**GlnValSer**GlyGluLeu 140
QY 374 GGAGGAGCAAGAACCTCCCGTGAAGCAAGGCAAGGCAAGGCTGATCTTGA 433

```

```

Db 140 gGluAspArganLeuProTyrSer**ArgAlaLysAlaLthrLeuLethrLeuPhe 160
QY 434 TACGAATACAGACCGTGAAGGGGGCTCAACGATCCTTCTGACCTTTGGGTTTA 493
Db 160 rThraSerThrAsp*****SerGlyAlaSerArgSerPhe**ProPheGlyPhe**Al 180
QY 494 AGGAGGTGCAAGAAAGTTACCAACAGGATTAAGTGGCTTGTGGCGGCAAGGTTCA 553
Db 180 eGly***ValArgLysValLthrThrGlyLethrGlyLeuThrPArgProSerVal**Se 200
QY 554 CGAGCTCGCTTTTGAATCCTTGATGCTGGCTCTTCTTCACTTGTGAAGCAATTA 613
Db 200 rAspValAlaPhe**SerPheAspValGlySerSerYrHis*****AlaGluPhe 220
QY 614 CAAGGTTGATGTTTACCCCACTAATAGGGAACGATGAGGATTAACCGTCTGAG 673
Db 220 rLysArgTrpLethrLthrLeuMet**CysCysCysHisGlyAsnProAlaGlnTyr 240
QY 674 ACAGGTAGTTTACCTTACCTGATGATGATGTTTGCATGATAATCTGCTCACTA 733
Db 240 GlnValSerPheThrLeuLeuMet**CysCysCysHisGlyAsnProAlaGlnTyr 260
QY 734 GAGGAACCGGAGGTTCAACATTTGTTGTAAGTGGCTGAGGAGGCAATGGGCGAA 793
Db 260 ValArgAsnArgArg*****HisLeuValLthrValLeuGly**GlyAlaAsnGly 280
QY 794 GCTACCATCTGTGGATTAATGACTGAACGCTTAACTAGTACAGATCCCGGAGG 853
Db 280 rLeu**SerValGlyLeu**LeuAsnAlaSerLysSerGlnSerArgProGly**Th 300
QY 854 GATACGCGAGCGCGCGGAGCTCGGCTGGCTCGGATAGCCGGTCCCGCTTCC 913
Db 300 rLethrGlnArgArgGlyAlaSerValGlyLeuGly**Pro*****ArgLeuSer 320
QY 914 GCGGCGGCGCGCGCGCGCGCTCAACGCGCGCGCGCGCGGAGGAGCGCGTCC 973
Db 320 oProAlaGlyArgProProLeuHisAlaProArg**ArgGlyArgAlaArgAlaPro 340
QY 974 GCGCGCGGAGCGCGGAGGTCGCGGAGTGCCTTCTGCTGCGGAAACGGGCGG 1033
Db 340 oArgAlaGlyLthrGlyValArgCys**ValProPheValLeuGlyAsnGlyAla 360
QY 1034 GGAAGGCGCGCGCGCGCGCTCGCGCTCAACGCAACCCAGTTCGTGGGAACCT 1093
Db 360 oGluArgArgPro*****ArgProSerArgThrAlaArgSerTrpGlyLthrTrp 380
QY 1094 AACCACTTCATCTCAGTCTCAGCTCGGCAAGCTGAGAGCCCTTCTCTCAAGA 1153
Db 380 *Thr**SerLethrSerProGlnProGlyLysLeuArgSerProPheLeuGln**G 400
QY 1154 GGTCAACCAACAGAGACCACTTTGGCAGAGAGCAGTGTGCATCTCAAGCCCG 1213
Db 400 nLeuThrGlnProGlnLthrHisPheGlyArgGluProAlaAla*****Ser 420
QY 1214 GGCAGATCTCCTGCTGAGAGCGCGCGCCAGCACTCTCATTTCTGTGCGAGCA 1273
Db 420 gAlaAspLeuProAlaGlnGluProAlaProSer**ProProCysLeuValGlnAla 440
QY 1274 AGAGGAGGCTGTGTGAGGAACCTCCAGAGAGAGACCTTCAAGAGCAGCCCA 1333
Db 440 uGlnGluAlaValLthrGlnLthrPro**GluGlnGlnLthrPheLysGlnLthr 460
QY 1334 GGTGCAACGCAAGGTGCTGTGAGCACTTGAACCAATTAACCAATTCAGAGGCG 1393
Db 460 uValGlnGlnGln*****GlySerGlnLthrHisLeuPheHisLethrGlnGly 480
QY 1394 CAGTGGCAAGGCGCTGTGCGCGCGCGCTTGAACGCTCAAGGAGCGAGCAAGCA 1453
Db 480 uSer**GlnGlyLeuCysAlaArgAlaLeuLysrAspLysGlnAlaAlaAsp 500
QY 1454 GATCTCTTTGAACCGGAGCACTCAACGCGGATCGAGGTATCGAAGAGCGTGTG 1513
Db 500 uLethrPheAspProGluAsnLeuLethrGlyLethrGlyVal*****GluGly 520

```

Fri Apr 2 07:35:03 2004

us-10-028-952a-4.rapb

Page 4

QY 1514 GCGTGGCTATGGCCGCGATGGCATTTTGGCATTTTCCCTGCCAACTAGTGGAGCTCAT 1573  
DB 520 pARGGLYTRGlyProAspGLYHisphedgylmec\*\*\*ProAlaasnTYrValGLIleu11 540  
QY 1574 TGAAGTAGGCTGAGGGGACATCTTGGCTTCCCTTCCCTTCCGACATAGGCTTCTTATTTGCTG 1633  
DB 540 eglu\*\*\*Gly\*\*\*GlyHisIleLeu\*\*\*PheProSerGlnTrpLeuProTYrCysTr 560  
QY 1634 GAGAGAGAGGCTGGAGATTGACATTCACACTCTTCCAGAAATAGAGAGAGAGAGAGAG 1693  
DB 560 pLYSARGARGPro\*\*\*\*\*HisSerAlaLeuPheGlnGlu\*\*\*AspProGln\*\*\*G1 580  
QY 1694 ATGAGGCTTCAAGGCTCCCTCCGCTTGGGCACTAGGCTGTCAACCCCAATAGAGCA 1753  
DB 580 Y\*\*\*\*LeuArgAlaProSerGlyLeuAlaAspSerAlaCysHisProLYScys\*\*\*As 600  
QY 1754 TGGCTGGTATTCACACATCTTCCGATCCCTCCGACCTCCGACCTCCGACAGCTTGAGCT 1813  
DB 600 nGlyLeuValIleProThrHisProSerCysIleProArgPro\*\*\*\*\*ThrAlaTrpLe 620  
QY 1814 CTGCCCCCTGACAGATTAAGTCAAGCCCTGCTGCTGAGCCAAAGCCCTGAGTGGCCACT 1873  
DB 620 uLeuProLeuThrGlyTYr\*\*\*AlaLYSProCys\*\*\*TrpProSerProGlnTrpProLe 640  
QY 1874 GCCAAGCTGGGGGAGAGGCTCCCTGAGCAGGAGGATCTGGGAGGCTGTGAGCTGCTTGC 1933  
DB 640 uProSerCysGlyGlnGlySer\*\*\*\*\*GlyAlaSerGlyArgLeuTrpLeuProSerAl 660  
QY 1934 ATTATTTGCTTTTCTTTTCTTTTCTTTGCTTAAAGGGTGGTGGCCACCACTGTTAG 1993  
DB 660 aPheIleCysLeu\*\*\*\*\*PheSerLeuAlaSerLYSGlyTrpTrpProProLeuPheArg 680  
QY 1994 AATGACCCCTGGGAGACAGTGAAGTGAAGATTGTTTAAAGCAGATTGTGACCAAGT 2053  
DB 680 gMet\*\*\*LeuGlyAsnSerGlnArgGlnLeuPheLeuAlaGlnPheValThr\*\*\*Va 700  
QY 2054 CAGAGTGAATCATGTTGGTGTGGCAGCAGGAAATTTGCTTGTGGAGCTGCTGTGTC 2113  
DB 700 lArgValAspHisGlyGlyLeuAlaIleGlyAsnLeuSerCys\*\*\*\*\*LeuLeuLYSAl 720  
QY 2114 TCCCATCTCATTTCTTCTTCTTCTTCTTCCCTGCGCTGAGGCTTGAAGTGGGATGGCAGTGGCC 2173  
DB 720 aProHisSerIleSerLeuSerLeuCysLeuGly\*\*\*GlyLYSArgGlyCysArgTrpTr 740  
QY 2174 AAGCTCCACCTGGTATTCATAAAGCGCAGACACAACTGTTCCTCCAGCGGAGCTCAC 2233  
DB 740 oSerSerHisProGlyTyrSerLYS\*\*\*AlaAspThrThrCysSerSerThrArgLeuTh 760  
QY 2234 TCGATGCTTCAAGGCCCCCAGTGTGTGCTCAACTGATTTGATTTGATTTGAGAAAGTAA 2293  
DB 760 rArgCysLeuGln\*\*\*\*\*ValCysAlaSerThrAspSerPheArgLYSLeuLYS 780  
QY 2294 AAAAAAAAAAACTCGAGAAGCTTTGAGCTTCTTCG 2330  
DB 780 sLYS\*\*\*LYSLeuLeuGlnLYSLeuTrpTrpHisSerSer 792

RESULT 2  
US-10-028-952a-10  
Sequence 10, Application US/10028952a  
Publication No. US20030157576A1  
GENERAL INFORMATION:  
APPLICANT: Eryn, Jr., Paul R.  
TITLE OF INVENTION: EPITHELIAL CELL GROWTH INHIBITORS  
FILE REFERENCE: 4273,3USM1  
CURRENT APPLICATION NUMBER: US/10/028, 952A  
PRIORITY FILING DATE: 2001-12-18  
PRIORITY FILING DATE: PCT/US00/16900  
PRIORITY FILING DATE: 2000-06-19  
PRIORITY FILING DATE: 1999-06-18  
PRIORITY FILING DATE: 1999-06-18  
NUMBER OF SEQ ID NOS: 10  
SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 10  
LENGTH: 784  
TYPE: PRN  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (110)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (13)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (16)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (33)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (49)..(50)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (66)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (83)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (90)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (99)..(100)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (104)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (116)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (133)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (142)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (147)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (149)..(150)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (166)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (173)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:

```

NAME/KEY: UNSURE
LOCATION: (183)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (199)..(200)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (204)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (216)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (233)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (242)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (249)..(250)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (255)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (266)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (281)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (283)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (299)..(300)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (316)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (333)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (349)..(350)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (366)..(368)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (378)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (382)..(383)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE

```

```

1 LOCATION: (386)
2 OTHER INFORMATION: Xaa is any amino acid or may be absent
3 FEATURE:
4 NAME/KEY: UNSURE
5 LOCATION: (398)..(401)
6 OTHER INFORMATION: Xaa is any amino acid or may be absent
7 FEATURE:
8 NAME/KEY: UNSURE
9 LOCATION: (416)
10 OTHER INFORMATION: Xaa is any amino acid or may be absent
11 FEATURE:
12 NAME/KEY: UNSURE
13 LOCATION: (427)
14 OTHER INFORMATION: Xaa is any amino acid or may be absent
15 FEATURE:
16 NAME/KEY: UNSURE
17 LOCATION: (433)
18 OTHER INFORMATION: Xaa is any amino acid or may be absent
19 FEATURE:
20 NAME/KEY: UNSURE
21 LOCATION: (449)..(450)
22 OTHER INFORMATION: Xaa is any amino acid or may be absent
23 FEATURE:
24 NAME/KEY: UNSURE
25 LOCATION: (456)
26 OTHER INFORMATION: Xaa is any amino acid or may be absent
27 FEATURE:
28 NAME/KEY: UNSURE
29 LOCATION: (466)
30 OTHER INFORMATION: Xaa is any amino acid or may be absent
31 FEATURE:
32 NAME/KEY: UNSURE
33 LOCATION: (443)
34 OTHER INFORMATION: Xaa is any amino acid or may be absent
35 FEATURE:
36 NAME/KEY: UNSURE
37 LOCATION: (499)..(500)
38 OTHER INFORMATION: Xaa is any amino acid or may be absent
39 FEATURE:
40 NAME/KEY: UNSURE
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
97
```

US-10-026-952A-4 (1-2333) X US-10-026-952A-10 (1-784)

QY 102 GCCCCCGCGCCCCCGGGTGTCCCGGAGGGGGCGGGGGGAGTCCGCGGCGCTGC 161  
Db 18 ALaProolValAProvalSerProArgValAlaArgValGly\*\*\*ArgArgProCys 37  
QY 162 GGCGCCGCGGTGAATACCACTACTGTGATCGTTTTTCATCTAGCCGGGTAGCGCGGGG 221  
Db 38 G1yProProValIySerTyhIstYzSerAspArg\*\*\*\*ThaSprProValAlaArgArgIly 57  
QY 222 GGAGAGCCCGGAGGGGCTCTCGCTTCTGTGGCGCAAGCCCGGCGCGCGCGCGCGGCG 281  
Db 58 G1yIuProArgG1yAlaLeuAla\*\*\*G1yAlaIyAspProAlaAlaArgArgProG1y 77  
QY 282 GGAGCCCGGCTCCGGGGAGACGTCCAGTGGGGAGTTTACTCGGGCGGTACACTGTCAA 341  
Db 78 AlAthrArgSerIy\*\*\*SerAlaArgItpolVal\*\*\*LeuG1yArgTyTthCysGln 97  
QY 342 ACCGTACGCAAGTGTCTTAAGCGAGACGTCCAGGAGGACAATAACCTCCGTGAGAGA 401  
Db 98 Thr\*\*\*\*\*GlnAlaSer\*\*G1yGluLeuArgGluAspAlaGluAsnProTyr\*\*\*Arg 117  
QY 402 AGGCGAAAAGCTCGTTGATCTTGAATTTCACTACGAATACAGACCGTGAAGCGGGGCC 461

```

Db      118 ArgAlaLysAlaArgLeuIleLeuIlePheSerThrAsnThrAsp***GluSerGlyAla 137
QY      462 TCACAGATCTCTTGAACCTTTGGGTTTAAGCAGAGAGGTGCAGAAAAGTTACACAGAGG 521
Db      138 SerArgSerPhe***ProPheGlyPhe***Ala*****ValArgLysValIleThrGly 157
QY      522 ATAAGTGGCTGTGGCGGCGCAAGCTTCATAGCAGAGCTGGCTTTTGAATCCTTCAGATGC 581
Db      158 IleThrGlyLeuThrPargProSer***HisSerAspValAlaPhe***SerPheAspVal 177
QY      582 GGCCTCTCTTATCATTTGTGAACAGAAATTCACCAAGCGTTGGATGTGTCAACCACTAATA 641
Db      178 GlySerSerThrHis***GluAlaGluPheThrLysArgTyrPileValHisProLeuIle 197
QY      642 GGAACGTGAGTGGGATTAAGACCGTCGTCGAGACAGTGAATTTACCTTACCTGATGATG 701
Db      198 Gly*****SerTyrVal***ThrValValArgIleValSerPheThrLeuLeu***Met 217
QY      702 TGTGTGTCATGATGATCTCTGCTCAATGTCAGAGAGAAACCGCAGGTTCAAGCATTTGGTG 761
Db      218 CysCysCysHisGlyAsnProAlaGlnTyrGluArgAsnArg***ArgHisLeuVal 237
QY      762 TATGTGCTTGGCTGAGGAGCAATGGGGGAGAGTACCATCTGTGGATTTGACTGATGAC 821
Db      238 TyrValLeuGly***GlyAlaAsnGlyAlaLys*****SerValGlyLeu***LeuAsn 257
QY      822 GCCTTAAGTCAAGATCCCGCCAGGCGGAAACGATACCGCAGCGCCGCGGAGCTCGATT 881
Db      258 AlaSerLysSerGlySerArgPro***GlyThrIleArgGlnArgGlyAlaSerVal 277
QY      882 GGCCTCGGATACCGCGGTCCCGCGCTGTCGCGCGGCGGCGGCGCGCC---CCCTCCAG 938
Db      278 GlyLeuGly***Pro***ProArgLeuSerProProAlaGlyArg-PropProPheSerTh 297
QY      939 GCGCCCGCGCGCGCGGAGGAGCGCGCTGCGCCCGCGCGCGCGGAGCGGGGTCCGGTGC 998
Db      297 Arg***ArgAlaGlyGlyArgValProArgAlaProGlyProGlySer***Al 317
QY      999 GGAGTGGCTTCTGCTCTGCGGAAACGGGCGCGCGCGGAAAGCGCGCCCTCGCGCCG 1058
Db      317 aglucysproserSerTyrGluThrGlyArgGlyArgLysGly***ProLeuAlaAr 337
QY      1059 TCACGACCGCAGCTTCGT 1077
Db      337 gHisAlaProHisValArg 343

RESULT 3
US-10-094-886-114
/ Sequence 114, Application US/10094886
/ Publication No. US20040002120A1
/ GENERAL INFORMATION:
/ APPLICANT: Kekuda, Ramesh
/ APPLICANT: Tchervnev, Velizar T.
/ APPLICANT: Liu, Xiaohong
/ APPLICANT: Spytek, Kimberly A.
/ APPLICANT: Patturajan, Meera
/ APPLICANT: Burgess, Catherine
/ APPLICANT: Vernet, Corinne A.
/ APPLICANT: Li, Li
/ APPLICANT: Gorman, Linda
/ APPLICANT: Malyanekar, Uriel M.
/ APPLICANT: Boldog, Ferenc
/ APPLICANT: Guo, Xiaojia
/ APPLICANT: Shenoy, Suresh
/ APPLICANT: Padigaru, Muralidhara
/ APPLICANT: Taupier, Raymond J., Jr.
/ APPLICANT: Miller, Charles
/ APPLICANT: Casman, Stacie
/ APPLICANT: Pena, Carol
/ APPLICANT: Gangoli, Esha
/ APPLICANT: Gusev, Vladimir
/ APPLICANT: Smithson, Glenda
/ APPLICANT: Zerhusen, Bryan

```

```

/ APPLICANT: Gerlach, Valerie
/ APPLICANT: Pochart, Pascal
/ APPLICANT: Fernandes, Elma
/ APPLICANT: Shimkets, Richard
/ APPLICANT: Rastelli, Luca
/ APPLICANT: Spaderna, Steven
/ APPLICANT: LaRocheille, William
/ APPLICANT: Zhong, Mei
/ TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD
/ FILE REFERENCE: 21402-290 B
/ CURRENT APPLICATION NUMBER: US/10/094, 886
/ CURRENT FILING DATE: 2002-03-07
/ PRIOR APPLICATION NUMBER: 60/274,322
/ PRIOR FILING DATE: 2001-03-08
/ PRIOR APPLICATION NUMBER: 60/313,182
/ PRIOR FILING DATE: 2001-08-17
/ PRIOR APPLICATION NUMBER: 60/288,052
/ PRIOR FILING DATE: 2001-05-02
/ PRIOR APPLICATION NUMBER: 60/318,510
/ PRIOR FILING DATE: 2001-09-10
/ PRIOR APPLICATION NUMBER: 60/274,281
/ PRIOR FILING DATE: 2001-03-08
/ PRIOR APPLICATION NUMBER: 60/314,018
/ PRIOR FILING DATE: 2001-08-21
/ PRIOR APPLICATION NUMBER: 60/274,194
/ PRIOR FILING DATE: 2001-03-08
/ PRIOR APPLICATION NUMBER: 60/274,849
/ PRIOR FILING DATE: 2001-03-09
/ PRIOR APPLICATION NUMBER: 60/296,693
/ PRIOR FILING DATE: 2001-06-07
/ PRIOR APPLICATION NUMBER: 60/313,626
/ PRIOR FILING DATE: 2001-08-21
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 298
/ SOFTWARE: PatentIn 2.1
/ SEQ ID NO: 114
/ LENGTH: 525
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-094-886-114

Alignment Scores:
Pred. No.: 3,57e-66 Length: 525
Score: 1191.00 Matches: 222
Percent Similarity: 99.55% Conservative: 0
Best Local Similarity: 99.55% Mismatches: 0
Query Match: 28.16% Indels: 1
DB: 15 Gaps: 0

US-10-028-952a-4 (1-2333) x US-10-094-886-114 (1-525)
QY      1095 ACCACCTTCATCTTCAGTCTCTCAGCTTCGACAGCTGAGAGCCCTTCTTCAGAGACAG 1154
Db      304 ThrThrSerIleSerSerProGlnProGlyLysLeuArgSerProPheLeuGlnLysGln 333
QY      1155 CTCACCCAGACAGAGACCCACTTTGGCAGAGAGCCAGCTGCTGCATCTCAAGGCGCAGG 1214
Db      324 LeuThrGlnProGlnThrHisPheGlyArgGluProAlaAlaAlaIleSerArgProArg 343
QY      1215 GCAGATCTCCCTGCTGAGAGAGCCGCGCCAGCAGCTCTTCATGTCTGTGTCAGGCAGAA 1274
Db      344 AlaAspLeuProAlaGluGluProAlaProSerThrProProCysLeuValGlnAlaGlu 363
QY      1275 GAGAGGCTGTGATATGAGAACTCCAGACAGAGACCTTTCAGAGAGAGCCCGCAGCTG 1334
Db      364 GluGluAlaValLysGluGluProProGluGlnGlnThrPheTyrGluGlnProProLeu 383
QY      1335 GTGCAGAGCAGAGGTGCTGCTGTCAGACATTGACCACTTGCAGGAGGAGGAGGCTC 1394
Db      384 ValGlnGlnGlnGlnLysGlnLysSerGluHisIleAspHisHisIleGlnGlnGlnLysLeu 403
QY      1395 AGTGGCAGAGGCTCTTGTGCGCCGCTTCAGACTTACAGGAGGCGAGCAGACAGAG 1454

```



Db	404	SerGlynginglyLeuGlySalAArgAlaLeuGlyAspGlyGlnAlaAlaAspAspArgGlu	423
QY	1555	ATCCCTTTTGAACCCCGAGAAACCTCATCAACGGGACATCGAGCGTGAATGACAGAAAGCGTGGTG	1514
Db	424	ILeserPheAspProGluAsnLeuLeuThrGlyIleGluValILesPheGluIlyTrpTrp	443
QY	1515	CGTGGCTATGGAGCCGAGATGGCCATTGTTGGCATGTTCCCTGCCAATCACTGAGAGCTCATTT	1574
Db	444	ArgGlyIlyTrGlyProAspGlyHisPhe-ALAcSerLeuProThrTrpTrpSerSerLe	463
QY	1575	GAGTAGGCTGAGGACATCTGTGCGCTTCCCGCTCGACAGATGGACTTCCTAATGCTGG	1634
Db	463	uSerGlnuIaGluGlyThrSerCysProSerProLeuArgHisGlyPheLeuIleAlaG	483
QY	1635	AAGAGGAGGCTGGGAGATTGACATTACAGACTCTTCCAGAAATAGACCCCGACATGAGAGA	1694
Db	483	YArgGlyIleuGlyValAspIleGlnHisSerSerArgAsnAlaGlyThrProSerGluAs	503
QY	1695	TGAGGCTCTCAGAGGCTCCCTCCGGCTGGCAGACTAGAGCTGTGCACCCCAATGACAGCAAT	1754
Db	503	pGluuIaSerGlyLeuProProAlaTrpGlnThrGlnProValThrProAsnAlaAlaLe	523
QY	1755	GGCTGTGG	1761
Db	523	CalatTrp	525

RESULT 4  
 US-09-879-957-192  
 Sequence 192. Application US/09879957  
 Patent No. US20020034755A1  
 GENERAL INFORMATION:  
 APPLICANT: SPARKS, Andrew B.  
 HOFFMAN, No. US20020034755A1n  
 KAY, Brian K.  
 FOWLER, Dana M.  
 MCCONNELL, Stephen J.  
 TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
 DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
 USING SAME  
 NUMBER OF SEQUENCES: 227  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Pennie & Edmonds LLP  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10036-2711  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0. Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/879,957  
 FILING DATE: 13-Jun-2001  
 CLASSIFICATION: <Unknown>  
 PRIORITY APPLICATION DATA:  
 APPLICATION NUMBER: US 08/630,915  
 FILING DATE: 03-APR-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Mastrock, S. Leslie  
 REGISTRATION NUMBER: 18,872  
 REFERENCE/DOCKET NUMBER: 1101-174  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 790-9090  
 TELEFAX: (212) 869-8864/9741  
 TELETYPE: 66141 PENNIE  
 INFORMATION FOR SEQ ID NO: 192:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 355 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: <Unknown>  
 TOPOLOGY: unknown

US-09-879-957-192	MOLECULE TYPE: Peptide	SEQUENCE DESCRIPTION: SEQ ID NO: 192:
Alignment Scores:	4.3e-66	Length: 355
Prod. No.:	1189.50	Matches: 251
Score:	73.28%	Conservative: 15
Percent Similarity:	69.15%	Mismatches: 63
Best Local Similarity:	28.12%	Indels: 34
Query Match:	9	Gaps: 6
US-10-028-952A-4 (1-2333) x US-09-879-957-192 (1-355)		
QY	724 CTCGATACGAGAGAACCGCAGGTTCAACATTTGGTGTATG-----TCGTTG	771
Db	12 LeuProGlyArgGlyThrProGlyProSerGlyLeuGlyValProGluAerGlnCysArg	31
QY	772 GCTAGGAGCCAAATGGAGCGCAAGTCAACCTGTGGGATATGACTGAAGCCTCTAAT	831
Db	32 ValArgAerLeuGlyTrpLeuAerSerPheThrAlaLysAlaGluLysGluGluGlu	51
QY	832 CAGAAATCCCGCCAGCGAGAACGATACGAGCGCCGCGAGAGCTCGATTGCTCGAT	891
Db	52 AsnArgArgLeuGluGluLysArgTrpAlaGluGluAlaGlnArgGlnLeuGluGlu	71
QY	892 AGCGGCTCCCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGC	951
Db	72 ArgArgGluArgGluLys-----ArgGluAlaAlaArgArgGluGlnArgTrpGlnGlu	89
QY	952 GCGGAGAGGC---GCGTCCCGCGCGCGCGC-----	979
Db	90 GlnGly-GlyGluLysAerProGlnSerArgTrpGluGlnGlnGlnValLys	109
QY	980 -CGGAGCCGGGGTCCGGAGCGAGATGCGCTTCGTCGGGAAACGGGCGCGCGGAAA	1038
Db	109 rArgAsnArgAsnGluGlnGlnSerAlaValHisProArgGluLys-PheGlyGlnLysG	129
QY	1039 GCGGCGCGCCCGCTCGCGCGCTCAGCAGCACCGCAGCTTGTGTGGGAACTGGCGCTMAACCA	1098
Db	129 LysGly-----AlaMetSer-----ThrT	135
QY	1099 CCTGCATCTCCAGTCTCCAGCGCTGCGCAAGCTGAGAGACCCCTTCGTCGAGAGAGCTCA	1158
Db	135 hrSerIleSerSerProGlnProGlyLysLeuArgSerProPheGlnGlnGlnGlnGln	155
QY	1159 CCGAACCAGAGACCCACTTTGGCAGAGAGCGAGTGTCTGCATCTTCAAGGCCACAGGCGAG	1218
Db	155 hrGlnProGluThrHisPheGlyArgGluProAlaAlaAlaIleSerArgProArgAla	175
QY	1219 ATTCCTCCGTGAGAGAGCGCGCGCCAGCACTCTCCATGTCTGTGAGCAGAGAGAGG	1278
Db	175 sPheProAlaGluGluProAlaProSerThrProProCysLeuValGlnAlaGlnGlu	195
QY	1279 AGCGTGTATAGAGAACCTTCAAGCAGAGAGACTTTCACGACAGCCCGCACTGTCG	1338
Db	195 LysAlaValArgGluGluProGluGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln	215
QY	1339 AGCAGCAAGAGTGTGCTGTGAGACATTTGACCAACCATTCAGAGGCGCGAGGCTCAATG	1398
Db	215 LngLngLngLngLysGlySerGlyLysHisIleAspHisIleGlnGlyLysGlyLeuSerG	235
QY	1399 GCGAAGGCGCTGTGCGCGCTGCTGTACGACTACAGGCAAGCCGAGACAGAGAGACT	1458
Db	235 LysGlnGlyLeuGlyAlaArgAlaGluLysArgTrpGlnAlaAlaAspAspThrGlnLys	255
QY	1459 CCTTTGACCCGCAAACTCATCAACGAGCATCGAGTGTGAGCAGAGAGCTGTGTCGTG	1518
Db	255 ePheAerProGluAsnLeuLysLeuThrGlyLysGlnValLysAspGlnGlyTrpArgG	275
QY	1519 GCTATGGGCGCGAGTGGCACTTTTGGCATGTTCCCTGCAACTAGTGGAGGCTCATTAAGT	1578
Db	275 LysTrpGlyProAspGlyHisPheGlyLysPheProAlaAsnTrpValGluLeuLysLeu	294

QY 1579 GAGGCTGAGGCGCATCTTCCCTCCCTCCAGACATGAGCTTCTTATGCTGGAAGA 1638  
| | | | |  
Db 295 GUAAGGUGGCTGTherCysProSerProLeuArgHisGlyPheLeuIleAlaGlyArg 314  
QY 1639 GAGGCGCTGGAGATTACATTCAGCACTCTTCCAGAAATAGACCCCGAGATGAG 1698  
| | | | |  
Db 315 GYGlyLeuGlyValAlaSerIleGlnHisSerSerArgAsnArgThrProSerGluAspGlu 334  
QY 1699 GCGTCAGGCGCTCCCTCCGCTGGAGCTCAGCTCAGCCCAATCAGCAATGAGCC 1758  
| | | | |  
Db 335 AlSerGlyLeuProProAlaThrGlnThrGlnProValThrProAsnIleAlaMetAla 354  
QY 1759 TGG 1761  
| | |  
Db 355 TTP 355  
RESULT 5  
US-10-424-599-256311  
; Sequence 256311, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated with  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21 (53223) B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ. ID NOS: 285664  
; SEQ. ID NO 256311  
; LENGTH: 270  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURES:  
; NAME/KEY: unsure  
; LOCATION: (1)..(270)  
; OTHER INFORMATION: unsure at all Xaa locations  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_73471C.1.pep  
US-10-424-599-256311  
Alignment Scores:  
Pred. No.: 8,93e-56 Length: 270  
Score: 1024.00 Matches: 223  
Percent Similarity: 82.42% Conservative: 2  
Best Local Similarity: 81.68% Mismatches: 29  
Query Match: 24.22% Indels: 21  
Gaps: 6  
US-10-028-952a-4 (1-2333) x US-10-424-599-256311 (1-270)  
QY 1142 GAGGCGGCTCTCTCAGCTTGCGAGCTGAGGACT-----GGAGAT 1104  
| | | | |  
Db 7 GUGGlyAla-----AlaLeuLeuArgThrLysPProArgProArgSerArgSer 23  
QY 1103 GAGAGTGTTTAAAGCGCAGGTTCCCAAGACGTGCGGTG---CGTACGCGGAGGAGG 1047  
| | | | |  
Db 24 ThrAsnGlyLeuAlaProGlySerThrArgThrCys-ValGlnArgAspGlyArgGluGlu 43  
QY 1046 CGGCGCGCTTCCGCGCGCGCGCGCTTCCAGAGAGAGGAGGCGGCGCGCGCGCGCG 987  
| | | | |  
Db 43 YGlyProLeuSerGlyArgThrProPheProLysThrAsnGlySerProHisArgThrPr 63  
QY 986 GATCCCG 927  
| | | | |  
Db 63 OValProThrProGlyYglYArg-Arg-----ArgAlaIleAlaGlyThrThra 80  
QY 926 GCGGCG 870  
| | | | |  
Db 80 rGlyProArgArgAlaGlyThrAspGlyAspProAlaIleArgGlyGlnProArgLeu 100

QY 869 G-CGGCGCTCCGATGCTTCCGCTGGCGGAGATTCTGACTTGAAGCGCTTGCATCA 811  
| | | | |  
Db 100 euaArgGysArgGlyValProProGlyArgAspSerAspLeuGluAlaPheSerHisA 120  
QY 810 ATCCCAAGATGATGCTTGGCCCATTTGGCTCTCTAGGCCAACAACATACCAAAATGC 751  
| | | | |  
Db 120 snProThrAspGlySerPheAlaProLeuAlaProGlnProSerThrTyThrLysCysL 140  
QY 750 TGAACCTGCGGTTCTCTCCACTG-AGCAGATTAACATGAGCAACACATCATCAT 692  
| | | | |  
Db 140 euaLeuAlaArgPheLeuSerTy\*\*SerArgIleThrMetAlaThrThrHisIleGln\* 160  
QY 691 AGGCTAAACTAATCTGCTTCAAGACGCTTAAATCCAGCTCAGCTTCTTATTAAGGG 632  
| | | | |  
Db 160 \*\*GlyLysThrAsnLeuSerHisAspGlyLeuAsnProAlaHisValProTy\*\*TrpV 180  
QY 631 TGAACAATCCACGCTTGTAATTCTGCTTCAATATGAGAAGAGCCGAGATCGAG 572  
| | | | |  
Db 180 alaSerAsnProThrLeuGlyGluPheCysPheThrMetIleGlyArgAlaAspIleGlu 200  
QY 571 GATCAAAAAGCGAGCTGCTATGAGCGCTTGCCGCGCAGCAGCTTATCCCTTGCTA 512  
| | | | |  
Db 200 LysSerLysSerAspValAlaMetAsnAlaTrpProProGlnAlaSerTyProCysAla 220  
QY 511 ACTTTCTGACACTCTCTGCTTAAACCAAAAGGTCA---GAAAGATCGTGAAGCCCG 455  
| | | | |  
Db 220 snPheSerAspThrSerCysLeuLysProLysArgSer\*\*Arg-AspArgGluAlaPro 239  
QY 454 CTTCACGCTGTATTGTAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 395  
| | | | |  
Db 240 LeuSerArgSerValPheValLeuLysIleLysIleLysArgAlaPheAlaLeuLeu\*\* 259  
QY 394 CACGAGAGCTTCTGCTCTCCCTCCGAGCTCGCT 362  
| | | | |  
Db 260 HisGlyArgPheLeuSer\*\*\*LeuSerSerPro 270  
RESULT 6  
US-09-879-957-18  
; Sequence 18, Application US/09879957  
; Patent No. US20020034755A1  
; GENERAL INFORMATION:  
; APPLICANT: SPARKS, Andrew B.  
; HOFFMAN, No. US20020034755A1h  
; KAY, Brian K.  
; FOWLES, Dana M.  
; MCCONNELL, Stephen J.  
; TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
; DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
; USING SAME  
NUMBER OF SEQUENCES: 227  
CORRESPONDENCE ADDRESS:  
ADDRESSSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/879,957  
FILING DATE: 13-Jun-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/630,915  
FILING DATE: 03-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Mirock, S. Leslie  
REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 1101-174  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 790-9090  
 TELEFAX: (212) 869-8864/9741  
 TELEX: 66141 PENNIE  
 INFORMATION FOR SEQ ID NO: 18:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 433 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: <Unknown>  
 TOPOLOGY: unknown  
 MOLECULE TYPE: peptide  
 SEQUENCE DESCRIPTION: SEQ ID NO: 18:  
 US-09-879-957-18  
 Alignment Scores:  
 Pred. No.: 1.07e-35 Length: 433  
 Score: 702.50 Matches: 146  
 Percent Similarity: 72.65% Conservative: 16  
 Best Local Similarity: 65.47% Mismatches: 35  
 Query Match: 16.61% Indels: 26  
 DB: 9 Gaps: 5  
 US-10-028-952a-4 (1-2333) x US-09-879-957-18 (1-433)  
 QY 971 GCCGCGCGCGGAGC-----CGGGGTCCGGTGGAGAGTCCCTGCTCGGAGAA 1021  
 Db 217 AlAlAlArgrgIugInArgrIYrGInGInGInHArSerAlaGlyAlaProSerArG 236  
 QY 1022 CGGGCGCGCGGAGAAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1081  
 Db 237 ThrGly-GluProGInGInGInu-----AlaValSerArgrIYrArGInGInuTrpG 253  
 QY 1082 AACG-----TGCGG 1090  
 Db 253 uSerAlaGlyGInGInAlaProHISProArGInuIlePheGInGInuGlyAlaGlyAlaHe 273  
 QY 1091 CTAACGACCTCCATCTCCAGTCCCTGAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAG 1150  
 Db 273 tSerThrThrSerValThrSerSerGInProGlyLysLeuArGSerProPheLeuGInu 293  
 QY 1151 GCGAGTCCGACCAAG 1210  
 Db 293 sGInLeuThrGInProGInuThrSerIYrGlyArGInuProThrAlaProValSerArGr 313  
 QY 1211 CAGGCGAGATCTCCCTGCTGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1270  
 Db 313 cAlaAlaGlyVal--CysGInuGInuProAlaProSerThrLeuSerSer--AlaGInuH 331  
 QY 1271 AGAAG 1330  
 Db 331 rGInuGInuGInuProThrIYrGInuValProProGInuGInuAspThrLeuIYrGInuProPr 351  
 QY 1331 ACTGATGACAG 1390  
 Db 351 cLeuValGInGInGInuGInuValaGlySerGInuHISLeaSPanTrYMeGInSerGInu 371  
 QY 1391 GCTCAGTGGGAG 1450  
 Db 371 yPheSerGlyGInuGInuLeuCYsaAlaGAlaLeuIYrAspIYrGInuAlaAlaAspAspH 391  
 QY 1451 AGAGATCTCCCTTACCCGAGAACTCATCAGCGGAGAGAGAGAGAGAGAGAGAGAGAG 1510  
 Db 391 rGInuIleSerPheAspProGInuHISLeuIleThrGInuValIleAspIYrGInuIYr 411  
 QY 1511 GTGGCGTGGCTATGAGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1570  
 Db 411 pTrpArGInuIYrGInuProAspGInuHISpHeGInuMerPheProAlaAsnIYrValGInu 431  
 QY 1571 CATGAG 1577  
 Db 431 uIleGInu 433

RESULT 7  
 US-09-925-299-1477  
 ; Sequence 1477, Application US/09925299  
 ; Patent No. US2002005627A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
 ; FILE REFERENCE: PA102  
 ; CURRENT FILING DATE: 2001-08-10  
 ; PRIOR APPLICATION NUMBER: PCT/US00/05883  
 ; PRIOR FILING DATE: 2000-03-08  
 ; PRIOR APPLICATION NUMBER: 60/124,270  
 ; NUMBER OF SEQ ID NOS: 1556  
 ; SOFTWARE: Patent Ver. 2.0  
 ; SEQ ID NO 1477  
 ; LENGTH: 90  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SITE  
 ; LOCATION: (53)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; NAME/KEY: SITE  
 ; LOCATION: (76)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; NAME/KEY: SITE  
 ; LOCATION: (87)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 ; US-09-925-299-1477  
 Alignment Scores:  
 Pred. No.: 1.29e-21 Length: 90  
 Score: 476.00 Matches: 87  
 Percent Similarity: 96.67% Conservative: 3  
 Best Local Similarity: 96.67% Mismatches: 0  
 Query Match: 11.26% Indels: 0  
 DB: 9 Gaps: 0  
 US-10-028-952a-4 (1-2333) x US-09-925-299-1477 (1-90)  
 QY 997 CACCGAGACCGCGGTCCCGGCGCGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 938  
 Db 1 HIsArGInuProValProAlaArGInuGlyAlaArGAlaLeuProArGAlaArGAla 20  
 QY 937 TGAAGGAGGAGGAGCGCGCGCGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 878  
 Db 21 TrpArGInuGlyArGProAlaGlyGlyAspArGArGInuGlyTrpArGrProArGrProthr 40  
 QY 877 GAGGTCGCGCGGCGGTGCGGTATGTTCCGCGCGGAGATTCGACTTGAAGCGGTC 818  
 Db 41 GAlaAlaProArGArGArGAlaGAlaProProGly\*\*AspSerAspLeuGInuAlaHe 60  
 QY 817 AGTCATAATCCACAGATGATGATGCTTCCGCCATGTGCTCTCCAGCAGATACAC 758  
 Db 61 SerHISaenProTruAspGlySerPheAlaProLeuAlaProGInu\*\*SerThrIYrThr 80  
 QY 757 AAATGTGGAACCTGCGGTCTCTCTCCGAC 728  
 Db 81 LysCYsaLeuAsnLeuArG\*\*LeuSerIYr 90  
 RESULT 8  
 US-09-925-299-1477  
 ; Sequence 1477, Application US/09925299  
 ; Publication No. US20030040617A9  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
 ; FILE REFERENCE: PA102  
 ; CURRENT FILING DATE: 2001-08-10  
 ; PRIOR APPLICATION NUMBER: PCT/US00/05883



;; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
;; FILE REFERENCE: PA102  
;; CURRENT APPLICATION NUMBER: US/09/925,299  
;; CURRENT FILING DATE: 2001-08-10  
;; PRIOR APPLICATION NUMBER: PCT/US00/05883  
;; PRIOR FILING DATE: 2000-03-08  
;; PRIOR APPLICATION NUMBER: 60/124,270  
;; PRIOR FILING DATE: 1999-03-12  
;; NUMBER OF SEQ ID NOS: 1556  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 1356  
;; LENGTH: 81  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: SITE  
;; LOCATION: (18)  
;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-299-1356

## Alignment Scores:

Pred. No.:	7,1e-19	Length:	81
Score:	432.00	Matches:	80
Percent Similarity:	98.77%	Conservative:	0
Best Local Similarity:	98.77%	Mismatches:	1
Query Match:	10.21%	Indels:	0
DB:	9	Gaps:	0

US-10-028-952a-4 (1-2333) x US-09-925-299-1356 (1-81)

QY 206 CCGGTTGAGCGGGGGGCGAGCCCGGAGGGGCTCTGCTTGTGGCCCAAGCCCGGC 265  
DB 1 ProGlyGluAlaGlyGlyArgAlaProArgGlySerThrPheTrpArgIn\*\*\*ProGly 20  
QY 266 CCGGCGCGCGCGCGGGCGGCGACCCGCTCCGGGAGACAGTGCAGGTGGGGATTGACTGGG 325  
DB 21 ArgAlaProAlaGlyArgAspProLeuArgGlyGlnCysGlnValGlySerLeuTrpGly 40  
QY 326 GCGGTACCTGTCAACGGTTAACGAGGTGCTTAAGCGAGCTCAGGAGACAGAAA 385  
DB 41 AlaValHisLeuSerAsnGlyAsnAlaGlyValLeuArgArgAlaGlnGlyGlnLys 60  
QY 386 CCGCCGTTGAGAGAGAGAGGCAAAAGCTCGCTGATCTTGAATTTTGAAGTACAGAA 445  
DB 61 ProProValGluGlnLysGlyLysSerSerLeuAspLeuAspPheGlnTrpGluTrpArg 80  
QY 446 CCG 448  
DB 81 Pro 81

## RESULT 11

US-09-925-299-1356  
;; Sequence 1356, Application US/09925299  
;; Publication No. US20030040617A9  
;; GENERAL INFORMATION:  
;; APPLICANT: Rosen et al.  
;; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
;; FILE REFERENCE: PA102  
;; CURRENT APPLICATION NUMBER: US/09/925,299  
;; CURRENT FILING DATE: 2001-08-10  
;; PRIOR APPLICATION NUMBER: PCT/US00/05883  
;; PRIOR FILING DATE: 2000-03-08  
;; PRIOR APPLICATION NUMBER: 60/124,270  
;; PRIOR FILING DATE: 1999-03-12  
;; NUMBER OF SEQ ID NOS: 1556  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 1356  
;; LENGTH: 81  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: SITE  
;; LOCATION: (18)

;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-299-1356

## Alignment Scores:

Pred. No.:	7,1e-19	Length:	81
Score:	432.00	Matches:	80
Percent Similarity:	98.77%	Conservative:	0
Best Local Similarity:	98.77%	Mismatches:	1
Query Match:	10.21%	Indels:	0
DB:	10	Gaps:	0

US-10-028-952a-4 (1-2333) x US-09-925-299-1356 (1-81)

QY 206 CCGGTTGAGCGGGGGGCGAGCCCGGAGGGGCTCTGCTTGTGGCCCAAGCCCGGC 265  
DB 1 ProGlyGluAlaGlyGlyArgAlaProArgGlySerThrPheTrpArgIn\*\*\*ProGly 20  
QY 266 CCGGCGCGCGCGCGGGCGGCGACCCGCTCCGGGAGACAGTGCAGGTGGGGATTGACTGGG 325  
DB 21 ArgAlaProAlaGlyArgAspProLeuArgGlyGlnCysGlnValGlySerLeuTrpGly 40  
QY 326 GCGGTACCTGTCAACGGTTAACGAGGTGCTTAAGCGAGCTCAGGAGACAGAAA 385  
DB 41 AlaValHisLeuSerAsnGlyAsnAlaGlyValLeuArgArgAlaGlnGlyGlnLys 60  
QY 386 CCGCCGTTGAGAGAGAGGCAAAAGCTCGCTGATCTTGAATTTTGAAGTACAGAA 445  
DB 61 ProProValGluGlnLysGlyLysSerSerLeuAspLeuAspPheGlnTrpGluTrpArg 80  
QY 446 CCG 448  
DB 81 Pro 81

## RESULT 12

US-09-925-299-1333  
;; Sequence 1333, Application US/09925299  
;; Patent No. US20020055627A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Rosen et al.  
;; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
;; FILE REFERENCE: PA102  
;; CURRENT APPLICATION NUMBER: US/09/925,299  
;; CURRENT FILING DATE: 2001-08-10  
;; PRIOR APPLICATION NUMBER: PCT/US00/05883  
;; PRIOR FILING DATE: 2000-03-08  
;; PRIOR APPLICATION NUMBER: 60/124,270  
;; PRIOR FILING DATE: 1999-03-12  
;; NUMBER OF SEQ ID NOS: 1556  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 1333  
;; LENGTH: 94  
;; TYPE: PRT  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: SITE  
;; LOCATION: (1)  
;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
;; NAME/KEY: SITE  
;; LOCATION: (15)  
;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
;; NAME/KEY: SITE  
;; LOCATION: (38)  
;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
;; NAME/KEY: SITE  
;; LOCATION: (42)  
;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
;; NAME/KEY: SITE  
;; LOCATION: (53)  
;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
;; NAME/KEY: SITE  
;; LOCATION: (88)  
;; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-299-1333



Fri Apr 2 07:35:03 2004

us-10-028-952a-4.rapb

Page 13

Db 21 LeuAlaSerAspSerArgSerProAlaCySProAlaArgAlaAlaProProSerThr 40  
QY 940 CGCCCCCGCCCGCGGAGGCGCCGTCGCCCCCGCGCGGAGACCGGGGTCCGGTGG 999  
Db 41 ArgProAlaArgAlaGlyArgValProArgAlaProGlyProGlySerGlyAla 60  
QY 1000 GAGTGCCCTTCGTCCTGGGAAACGGGGCGCGCGGAAAGCGCGCGC 1047  
Db 61 GlucCyProSerSerTrpGluThrGlyProGlyTrpLysGlyGlyArg 76

RESULT 15

US-09-925-1283  
; Sequence 1283, Application US/09925299  
; Publication No. US20030040617A9  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: F4102  
; CURRENT APPLICATION NUMBER: US/09/925,299  
; CURRENT FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05883  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 156  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1283  
; LENGTH: 91  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (88)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (91)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-299-1283

Alignment Scores:  
Pred. No.: 1,85e-16 Length: 91  
Score: 393.00 Matches: 74  
Percent Similarity: 97.37% Conservative: 0  
Best Local Similarity: 97.37% Mismatches: 2  
Query Match: 9.29% Indels: 0  
DB: 10 Gaps: 0

US-10-028-952a-4 (1-2333) x US-09-925-299-1283 (1-91)

QY 820 ACGCCTTAAGTCCGAAATCCCGCCAGCGGAGACGATACGCGACCGCGGAGCCTCGG 879  
Db 1 ThrProLeuSerGlnAsnProAlaGlnAlaGluArgTyrGlySerAlaGluProArg 20  
QY 880 TTGGCCTTCGAAATACCGCGGTCCCGCTGTCCCGCGCGGCGCGCCCTCCACG 939  
Db 21 LeuAlaSerAspSerArgSerProAlaCySProAlaArgAlaAlaProProSerThr 40  
QY 940 CGCCCCCGCCCGCGGAGGCGCGTGCCTCCCGCGCGCGCGGAGACCGGGGTCCGGTGC 999  
Db 41 ArgProAlaArgAlaGlyArgValProArgAlaProGlyProGlySerGlyAla 60  
QY 1000 GAGTGCCCTTCGTCCTGGGAAACGGGGCGCGCGGAAAGCGCGCGC 1047  
Db 61 GlucCyProSerSerTrpGluThrGlyProGlyTrpLysGlyGlyArg 76

Search completed: April 1, 2004, 16:17:15  
Job time: 111.5 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Comugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 1, 2004, 15:46:31 ; Search time 108.5 Seconds

(without alignments)  
12150.845 Million cell updates/sec

Title: US-10-028-952A-4

Perfect score: 4230

Sequence: 1 gacagagattcccatcttc.....agcttgcactcttcgcga 2333

Scoring table: BLOSUM62

Xgapop 10.0 , Xgapext 0.5

Ygapop 10.0 , Ygapext 0.5

Egapop 6.0 , Fgapext 7.0

Delop 6.0 , Delext 7.0

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 3172214

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_n2p\_model -DEV=xid  
-Q=/cgn2\_1/USPTO\_epool\_p/US10028992/runat\_01042004\_154624\_634/app\_query.fasta\_1.2503  
-DB=A.Geneseq\_29Jan04 -QFMT=faelan -SUFFIX=rag -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=bloms62 -TRANS=human40.cd1  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pcr -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=plc -NOEM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USPR=US10028952 @cgn 1.1.83 @runat 01042004\_154624\_634 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEIOBERY -NEG\_SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOC  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: A.Geneseq\_29Jan04:\*

1: geneseqp1980s:\*

2: geneseqp1980s:\*

3: geneseqp2000s:\*

4: geneseqp2001s:\*

5: geneseqp2002s:\*

6: geneseqp2003as:\*

7: geneseqp2003bs:\*

8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	3389	80.1	760	4	AAB66394 Human pro
2	1370.5	32.4	753	4	AAB66395 Human mam
3	1191	28.2	525	6	ABUS5367 Human GPC
4	1189.5	28.1	355	2	AAW05398 Human CLO
5	1033	24.4	538	2	AAV23756 Human mam
6	872	20.6	302	5	ABR04724 Human pps
7	872	20.6	430	4	AAW78585 Human pro
8	870	20.6	488	4	AAW78569 Human pro
9	872	20.6	199	3	AAB65214 Human pro
10	870	20.6	203	5	ABB57455 Human sec

11	870	20.6	217	2	AAV73941 Human pro
12	870	20.6	377	6	ABP98850 Human str
13	870	20.6	431	3	AAB20896 Human dre
14	870	20.6	431	3	AAV85662 Human tyr
15	870	20.6	439	4	AAB93895 Human pro
16	716	16.9	162	2	AAV73942 Human pro
17	702.5	16.6	433	2	AAW05389 Mouse SH3
18	695	16.4	201	3	AAB63213 Gene 38 h
19	476	11.3	90	3	AAB53937 Human col
20	472.5	11.2	118	4	AAV7066 Human col
21	447.5	10.6	142	4	ABG14272 Novel hum
22	432	10.2	81	3	AAB53816 Human col
23	430	10.2	80	5	ABBS7422 Human sec
24	430	10.2	80	6	ABP99731 Human sec
25	430	10.2	80	6	ABR01218 Human gen
26	430	10.2	80	6	ADA98322 Human sec
27	430	10.2	80	6	ADA44146 Human sec
28	430	10.2	80	7	ADCC20492 Human sec
29	430	10.2	81	3	AAB63171 Human col
30	405	9.6	94	3	AAB53793 Human col
31	393	9.3	91	3	AAB53743 Human col
32	374.5	8.9	227	4	AAU33240 Novel hum
33	373	8.8	73	3	AAB53817 Human col
34	345	8.2	74	3	AAB44205 Human can
35	332	7.8	74	3	AAB53862 Human col
36	324	7.7	58	2	AAW34223 SH3 domai
37	324	7.7	58	2	AAW34248 SH3 domai
38	319	7.5	85	4	ABG00727 Novel hum
39	309	7.3	59	4	AAW95597 Human rep
40	308	7.3	301	4	AAW83949 Human imm
41	277	6.5	486	6	ABR44110 Human pro
42	277	6.5	486	6	ABO07214 Human p53
43	277	6.5	486	7	ADD67624 Human Ly1
44	267	6.3	54	3	AAW53774 Human col
45	266	6.3	54	4	AAW96009 Human rep

## ALIGNMENTS

### RESULT 1

AAB66394 standard: protein; 760 AA.

ID AAB66394;

AC AAB66394;

DT 10-APR-2001 (first entry)

DE Human prostate ECGI protein sequence.

KW Human; mammasatatin A; mammasatatin B; mammasatatin C; ECGI; cancer;

OS Homo sapiens.

PN WO200078955-A1.

PD 28-DEC-2000.

PF 19-JUN-2000; 2000WO-US016900.

PR 18-JUN-1999; 99US-0139995P.

PA (BIOT-) BIOTHERAPIES INC.

PI Ertvin PR;

DR WPI; 2001-07193/08.

DR N-PSDB; AAF31282.

PT A Mammasatatin-like epithelial cell growth inhibitor, and the nucleic that  
encodes it, useful for diagnosing and/or preventing epithelial cell  
cancers, e.g. of the ovaries or prostate.

PS Disclosure: Page 41-43; 55pp; English.

XX The present invention describes a mammostatin-like epithelial cell growth  
 CC inhibitor (ECGI) which has substantial similarity to mammostatin A, B or  
 CC C. ECGI is expressed in healthy cells, but is either absent or expressed  
 CC at reduced levels in cancerous cells. The protein and its coding sequence  
 CC can be used to inhibit epithelial cell growth and the amount present in  
 CC cells can be used to diagnose cancer or monitor its progress

XX Sequence 760 AA;

Alignment Scores:

Pred. No.:	4,13e-272	Length:	760
Score:	3389.00	Matches:	695
Percent Similarity:	87.64%	Conservative:	0
Best Local Similarity:	87.64%	Mismatches:	49
Query Match:	80.12%	Indels:	49
DB:	4	Gaps:	31

US-10-028-952A-4 (1-2333) x AAB66394 (1-760)

```

QY 2 CACGAGATTCCCACTGTCCCTTACTACTATCCAGGAAACCAACAGCGAAACGGGCT 61
Db 1 HtSGuileProthThValProthTyTyProalalysProGln****GluArgAla 20
QY 62 TGGCGGAATCAGCGGGGAAAGAGACCTGTGAGC----- 97
Db 21 TTPARGaEnGlnARgGlyLysLysThLeuLeuSerThLeuValTTPHIGlyGlu 40
QY 98 -----TTGGGCCCCCGCGCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGAG 133
Db 41 Thr***GluVal***AsnLysTrp***AlaProGlyAlaProProValSerProArgG 60
QY 134 GGGCGCGGGCGGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 193
Db 60 YAlaArgGlyGly-----ArgProCysGlyProProValLysTyRHSYRSeAspAr 79
QY 194 TTTTCTACTGACCGCGGTAGGCGGGGGGCGAGCGCCGAGGGGCGCTCTCGTTCTGCGC 253
Db 79 g---PhetThrAspProValARgARgGlyGlyGluProARgGlyAlaLeuAlaSer***Al 98
QY 254 CAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 313
Db 98 alysaARgProAlaAlaARgARgProGlyAlaThrxysersGly-----AlaARgTrpG 117
QY 314 AGTTGACTGGGGCGGTACACCTGTCAACCGTAACGAGAGCGAGTGTCTTAAGCGAGCTCAG 373
Db 117 yVal***LeuGlyARgTyThrCysGlnThrVal---GlnValSer***GlyGluLeuAR 136
QY 374 GAGAGACAGAAACCTCCCGGTAGACAGAGAGGCGAAAGCTCGCTGATCTGATTTTCA 433
Db 136 gGluAspARgAsnLeuProTrpSer***ArgAlaLysAlaARgLeuLeuLeuLeuLeuPheSe 156
QY 434 TAGAATTAAGACCGGAAAGCGGGCGCTCAAGATCTCTGACCTTTGGTTTAAAGC 493
Db 156 rThraAsnTrpAsp-----SerGlyAlaSerARgSerPhe***ProPheGlyPhe----* 174
QY 494 AGAGAGTGTCAAAAAGTTAACCAAGAGATAACTGTGCTTGCGCGCGCAAGCTTCATAG 553
Db 174 *AlaGlyValARgTyValThrThrGlyLeuThrPARGProSerVal**Se 194
QY 554 CGAGCGCGCTTTTGAATCCCTGAGTGGCGCTTCTCATTTGAGAGAGCAATTCAC 613
Db 194 rAspValAlaPhe***SerPheAspValGlySerSerTyRHS-----AlaGluPheH 213
QY 614 CAAGCGTGTGATTTTCAACCACTTAATAGAGAACTGAGCTGGAGTTAGACCTGCTGAG 673
Db 213 rLysARgTrpIlaValARgProLeuIleGlyAsn---SerTrpAsp***ThValValAR 232
QY 674 ACAGGTAGCTTAACTCTAGTGAATGTGTTGTCAGATGGAATCTCGCTCAGAGACA 733
Db 232 gGlnValSerPheThrLeuLeuMet***CysCysGlySHIGlyAsnProAlaGlnTyG 252

```

```

QY 734 GAGAAACCGCAGGTTCCAGACATTTGATGTGCTTGAGTGAAGACCAATGGGGCGAA 793
Db 252 uArgaAsnARgARg-----HtLeuValTyRValLeuGly***GlyAlaAsnGlyAlaTy 271
QY 794 GCTACCATCTGGGGATTATGACTGAAGCGCTCTTAAGTACAGAAATCCCGCCAGCGGAAC 853
Db 271 sLeu---SerValGlyLeu***LeuAsnAlaSerTySerSerGlySerARgProGly***Th 290
QY 854 GATACGACAGCGCGCGGAGACCTCGATTGCGCTCGAATACCGCGTCCCGCTGTCCCC 913
Db 290 rLleARgGlnARgARgGlyAlaSerValGlyLeuGly***Pro****ArgLeuSerPr 309
QY 914 GCGCGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 973
Db 309 oProAlaGlyARgProProLeuHtAlaProARg---ARgAlaYARgAlaARgAlaProPr 328
QY 974 GCGCGCGGGAGCCGGGGTCCCGGTGCGAGAGTGCCTTCGTCGGAAGAACGGGCGCGCC 1033
Db 328 oARgAlaGlyThrGlyValARgCys***ValProPheValLeuGlyARgGlyAlaARgPr 348
QY 1034 GGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1093
Db 348 oGluARgARg---Pro***ArgProSerARgThrAlaARgSerTrpGlyThrTrpARg-- 366
QY 1094 AACCACTCCATCTCCAGTCTCCAGCTCGCAAGCTGAGAGAGCCCTTCCTGCAGAGACA 1153
Db 367 ---TherLleSerSerProGlnProGlyLysLeuARgSerProPheLeuGln***G 385
QY 1154 GCTCACCCAAACAGAGAACCACTTTGGAGAGAGAGCGAGTGTCTGCATCTCAAGGCCAG 1213
Db 385 nLeuThrGlnProGlnThrHtAspHeGlyARgGluProAlaAla***-SerARgProAR 404
QY 1214 GCGAGATCTCTCTGTGAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1273
Db 404 gAlaAspLeuProAlaGluGluProAlaProSer---ProProCysLeuValGlnAlaG 423
QY 1274 AGAGAGGCTGTGTATGAGAGAACCTCCAGAGCAGAGACCTTCTACGACAGCCCACT 1333
Db 423 uGluGluAlaValTyGluGluPro**GluGlnGlnThrPheTyGluGlnProProLe 443
QY 1334 GGTGCGACAGCAAGGTGCTGCTGAGCAGCATTTGACCAACCAATTCAGGGCGCGGCT 1393
Db 443 uValGlnGlnGln---GlySerGlnHtSleARgPheSHISleGlnGlyGlnLysLe 462
QY 1394 CAGTGGCAAGGAGCTCTGTGCGCGCGCTGTACGACTACAGCGAGCGGAGACACACA 1453
Db 462 uSer---GlnGlyLeuCySAlaARgAlaLeuTyRAspTyRAlaAlaAlaAspAsp**G 481
QY 1454 GATCTCTTGAACCCCGAGAACTCATCAGGGGATCGAGGTGATGAGACGAGGCTGTG 1513
Db 481 uLleSerPheAspProGluAsnLeuLeuThrGlyLleGluVal-----GluGlyTrpTr 500
QY 1514 GCGGTGATGAGCGCGGATGGCAATTTGGCATGTTCCCTGCCAATCTAGTGAAGTCA 1573
Db 500 PARgGlyTyRgTyProAspGlyHtAspHeGlyMet---ProAlaAsnTyRValGluLeu 519
QY 1574 TGGTGAAGCTGAGGGAGACATCTTGCCCTTCCCTGACAGACATGAGTTCCTATTGCTG 1633
Db 519 eGlu***Gly***GlyHtSleLeu***PheProSerGlnThrTrpLeuProTyRysTr 539
QY 1634 GAAAGAGAGCGCTGGAGTTGACATTCAGCACTTTCAGAGAAATGAGACCCCAAGTGA 1693
Db 539 pLysARgARgPro*****HisSerAlaLeuPheGlnGlu***AspProGln***G 558
QY 1694 ATGAGGCTCAGGGCTCCCTCGGCTTGGCAGACTAGACTGTCAACCCCAATATGACAA 1753
Db 558 y---LeuARgAlaProSerGlyLeuAlaAspSerAlaCysHisProTyRys***As 577
QY 1754 TGGCGTGTGATTTCCACACATCTTCTGCAATCCCGGACCTCCCAAGACAGCTTGCT 1813
Db 577 nGlyLeuValIleProThrHtAspSerCysIleProARgPro-----ThrAlaTrpLe 595
QY 1814 CTTCGCCCTGACAGATACGAGACCAAGCCCTGCTGTGCGCAAGCCCTGAGTGGCCACT 1873

```

```

|||||
Db 595 uleuProLeuThGlyTyr***AlaLysProCys---TPrProSerProGluTrpProLe 614
1874 GCCAAGCTCGGGGAGGAGCTCTGAGCAGGAGCATCTGGAGAGCTGGCTCTGCTGC 1933
614 uProSerCysGlyGluGlySer*****GlyAlaSerGlyArgLeuTrpLeuProSerAl 634
QY 1934 ATTATTTGCTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTG 1993
634 apHeileCysLeu---PheSerLeuAlaSerLysGlyTrpTrpProLeuPheAr 653
QY 1994 AATGACCTTGGGAGACGTAGACGTAGAGATTTGTTTGAAGAGATTGTGTGACCAAAGT 2053
653 gMet---LeuGlyAsnSerGluArgArgGluLeuPheLeuAlaGluPheValThr***Va 672
QY 2054 CAGAGTGGATCATGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGC 2113
672 laGValAspHisGlyGlyLeuAlaAlaGlyAsnLeuSerCys---**LeuLeuCysAl 691
QY 2114 TCCCACTCATTTCTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGTGC 2173
691 apProHisSerLysSerLeuSerLeuSerLeuGly---GlyLysTrpGlyCysArgTrpPr 710
QY 2174 AAGCTCCACCTCTGGGTATTCAAAACGGCAGACACACATGTTCTCTCCAGCGGCTCAC 2233
710 oSerSerHisProGlyTyrSerLys**AlaAspTrpTrpCysSerSerThrArgLeuTh 730
QY 2234 TCAGATGCTTCAGACCCCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2293
730 raGcysLeuGln***-ValCysAlaSerThrAspSerAspPheArgLysSer---Ly 748
QY 2294 AAAAAAAAAAACTCGAGAGAGCTTGGACTTCTTGC 2330
748 sLysLysLysLeuGluLysLeuTrpTrpSerSer 760
Db

```

```

RESULT 2
AAB6395
ID AAB6395 standard; protein; 753 AA.
XX
AC AAB6395;
XX
DT 10-APR-2001 (first entry)
XX
DE Human mammaplastin C protein sequence.
XX
KM Human; mammaplastin A; mammaplastin B; mammaplastin C; ECGI; cancer;
XX
OS Homo sapiens.
XX
PN W0200078955-A1.
XX
PD 28-DEC-2000.
XX
PF 19-JUN-2000; 2000WC-US016900.
XX
PR 18-JUN-1999; 99US-0139995P.
XX
PA (BIOT-) BIOTHERAPIES INC.
XX
PI Ervin PR;
XX
DR WPI; 2001-071393/08.
XX
PS
XX
PT A Mammaplastin-like epithelial cell growth inhibitor, and the nucleic that
XX
CC encodes it, useful for diagnosing and/or preventing epithelial cell
XX
PT cancers, e.g. of the ovaries or prostate.
XX
PS Disclosure; Page 44-46; 55pp; English.
XX
CC The present invention describes a mammaplastin-like epithelial cell growth
XX
CC inhibitor (ECGI) which has substantial similarity to mammaplastin A, B or
XX
CC C. ECGI is expressed in healthy cells, but is either absent or expressed

```

CC at reduced levels in cancerous cells. The protein and its coding sequence  
 CC can be used to inhibit epithelial cell growth and the amount present in  
 CC cells can be used to diagnose cancer or monitor its progress

Sequence 753 AA:

Alignment Scores:

Pred. No.:	1,14e-104	Length:	753
Score:	1370.50	Matches:	286
Percent Similarity:	88.04%	Conservative:	1
Best Local Similarity:	87.73%	Mismatches:	27
Query Match:	32.40%	Indels:	12
DB:	4	Gaps:	11

US-10-028-952a-4 (1-233) x AAB6395 (1-753)

```

QY 102 GCCCCCGGCGCCCGCGGTGTCCCGCAGAGGGCCCGGGCGGGTCCGCGGCTCTGC 161
Db 17 AlaProGlyAlaProProValSerProArgGlyAlaArgGlyGly***ArgArgProCys 36
QY 162 GGGCGCGCGGTGAATACCACTACTGATGTTTTCACCTGACCCGGTGAAGCGGGGG 221
Db 37 GlyProProValLysTyrHisTyrSerAspArg---**ThrAspProValArgArgGly 55
QY 222 GGGCGAGCCCGAGGAGGCTCTGCTTGTGCGCGCAGAGCGCCGCGCGCGCGCGCGGC 281
Db 56 GlyLysProArgGlyAlaLeuAla---GlyAlaLysArgProAlaAlaArgArgProGly 74
QY 282 GCGACCCGCTCCGGGACAGTCCAGAGTGGAGAGTGTGATCGGGCGGTACCTGTGCA 341
Db 75 AlaThrArgSerGly***SerAlaArgTrpGlyAla***LeuGlyArgGlyTrpCysGln 94
QY 342 ACGGTAACGAGGTGTCTTAAGCGAGCTCAGAGAGACAGAAACCTCCGTGAGAGAGA 401
Db 95 Thr---**GlnValSer***GlyGluLeuArgLysAspArgAsnLeuProTrp---Arg 112
QY 402 AGGCGAAAGCTCGCTTATCTTGAATTTTCATGATGATACAGACCGTGAAGCGGGGCC 461
Db 113 ArgAlaLysAlaArgLeuLeuLeuPheSerThrAsnThrAsp***GluSerGlyAla 132
QY 462 TCAGATCCTCTGACCTTTGGGTTTAAGCAGAGAGTGCAGAAAAGTTACACAGAGG 521
Db 133 SerArgSerPhe***ProPheGlyPhe---**Ala***ValArgLysValThrThrGly 151
QY 522 ATAAGTGGCTTGTGGCGGCGCAGAGCTTCATAGCAGCGTCTTTTGAATCTTGATGTC 581
Db 152 IleThrGlyLeuTrpArgProSer---HisSerAspValAlaPhe***SerPheAspVal 170
QY 582 GGGCTTCTCTATCTTGTGAAGCAGAAATTCACAGAGGTTGATGTTACCCACTATA 641
Db 171 GlySerSerTyrHis***GluAlaGluPheThrLysArgTrpIleValHisProLeuIle 190
QY 642 GGAACGTGAGCTGGAGATTAGACCGTGTGAGACAGAGTTAGTTTACCTTACATGATG 701
Db 191 Gly---**SerTrpVal**ThrValValArgGlnValSerPheTrpLeuLeu---Met 208
QY 702 TGTGTTGCTATGTAATCCTGCTCAGTACGAGAGGAAACCGCAGGTTACAGATTGSGTG 761
Db 209 CysCysCysHisGlyAsnProAlaGlnTyrGluArgAsnArgArg***ArgHisLeuVal 228
QY 762 TATGTGCTTGGCTGAGAGCAATGGGGCGAAGCTTACCATCTGTGGATTATGACTGAAC 821
Db 229 TyrValLeuGly***GlyAlaAsnGlyAlaLys***-SerValGlyLeu***LeuAsn 247
QY 822 GCGCTTAAGTACGAATCCCGCGCAGCGCAGACGATACGCGCGCGCGGAGCGCTCGGAT 881
Db 248 AlaSerLysSerGluSerArgPro---GlyThrIleArgGlnAspArgGlyAlaSerVal 266
QY 882 GGCCTCGGATAGCCGGTCCCGGCTGTCCCGCGGCGGCGCGCGCCCTCCACAGCG 941
Db 267 GlyLeuGly***Pro***ProArgLysSerProProAlaGlyArg-PropProProSerTh 286
QY 942 CCGCGCGCGCGGAGGCGCGTCCCGCGCGCGCGCGGAGCCGCGGCTCGGTCGCGA 1001

```

```
Db      286 rArg**AgaAlaGlyArgValProArgAlaProGlyProGlySer---AlaGI 305
QY      1002 GTGCCCTTCCTCGTGGAAACGGGCGCGCCGGAAGCGCGCCCTCGCCCTCA 1061
Db      305 ucysproserisertirglutThrGlyArgGlyArgIysglYgIy***ProLeuAlaArgHt 325
QY      1062 CGCACCGCACGCTTCCT 1077
Db      325 sAlaProHisValArg 330

RESULT 3
ABUS2367
ID      ABUS2367 standard; protein; 525 AA.
XX
AC      ABUS2367;
XX
DT      03-MAR-2003 (first entry)
XX
DE      Human GPCR related protein NOV22a.
XX
KW      Human; NOVX; G-protein coupled receptor; GPCR; cancer; cyrostatic.
XX
OS      Homo sapiens.
XX
PN      WO200279398-A2.
XX
PD      10-OCT-2002.
XX
PF      08-MAR-2002; 2002WO-US007355.
XX
PR      08-MAR-2001; 2001US-0274194P.
PR      08-MAR-2001; 2001US-0274281P.
PR      08-MAR-2001; 2001US-0274322P.
PR      09-MAR-2001; 2001US-0274849P.
PR      13-MAR-2001; 2001US-0275578P.
PR      13-MAR-2001; 2001US-0275579P.
PR      13-MAR-2001; 2001US-0275601P.
PR      14-MAR-2001; 2001US-0276000P.
PR      16-MAR-2001; 2001US-0276776P.
PR      19-MAR-2001; 2001US-0276994P.
PR      20-MAR-2001; 2001US-0277239P.
PR      20-MAR-2001; 2001US-0277327P.
PR      20-MAR-2001; 2001US-0277338P.
PR      21-MAR-2001; 2001US-0277791P.
PR      22-MAR-2001; 2001US-0277833P.
PR      23-MAR-2001; 2001US-0278152P.
PR      26-MAR-2001; 2001US-0278894P.
PR      27-MAR-2001; 2001US-0278899P.
PR      27-MAR-2001; 2001US-0279036P.
PR      30-MAR-2001; 2001US-0280233P.
PR      02-APR-2001; 2001US-0280802P.
PR      02-MAY-2001; 2001US-0280532P.
PR      02-MAY-2001; 2001US-0280566P.
PR      17-MAY-2001; 2001US-0288228P.
PR      07-JUN-2001; 2001US-0291766P.
PR      08-JUN-2001; 2001US-0296653P.
PR      08-JUN-2001; 2001US-0296856P.
PR      05-JUL-2001; 2001US-0303230P.
PR      05-JUL-2001; 2001US-0303237P.
PR      08-AUG-2001; 2001US-0310913P.
PR      13-AUG-2001; 2001US-0311978P.
PR      14-AUG-2001; 2001US-0312191P.
PR      16-AUG-2001; 2001US-0312916P.
PR      17-AUG-2001; 2001US-0313182P.
PR      20-AUG-2001; 2001US-0313408P.
PR      21-AUG-2001; 2001US-0314018P.
PR      27-AUG-2001; 2001US-0315227P.
PR      10-SEP-2001; 2001US-0318403P.
PR      10-SEP-2001; 2001US-0318510P.
PR      14-SEP-2001; 2001US-0322296P.
PR      14-SEP-2001; 2001US-0322360P.
PR      27-SEP-2001; 2001US-0325378P.
```

```
PR      09-NOV-2001; 2001US-0332486P.
PR      09-NOV-2001; 2001US-0345399P.
PR      07-MAR-2002; 2002US-00094886.
XX
XX      (CURA-) CURAGEN CORP.
XX
PI      Kekuda R, Tchervnev VN, Liu X, Spytek KA, Paturajan M;
PI      Burgess CE, Vernet CM, Li L, Gorman L, Malvankar UM, Boidog FL;
PI      Guo X, Shenoy SG, Padigaru M, Taupier RJ, Miller CE, Casman SJ;
PI      Pena CE, Gangoli EA, Gusev V, Smithson G, Zethusen BD, Gerlach V;
PI      Pochart PF, Fernandes ER, Shinkets RA, Rastelli L, Spaderna SK;
PI      Larochele WJ, Zhong M, Kiramtsov NV, Voss EZ, Herrmann JU;
XX
XX      WPI; 2003-058423/05.
DR      N-PSDB; ABX70458.
XX
XX      NOVX polypeptides and polynucleotides, useful for treating a syndrome
XX      related to a human disease associated with the NOVX polypeptide e.g.,
XX      cancer.
XX
XX      Claim 1; Page 156; 413pp; English.
XX
CC      The present invention relates to the isolation of novel human
CC      polypeptides referred to as NOVX (NOV1-NOV44), variants of these
CC      proteins, and the polynucleotide sequences encoding them. The NOVX
CC      proteins of the invention are G-protein coupled receptor (GPCR) related
CC      proteins. The sequences of the invention are useful in the manufacture of
CC      a medicament for treating a syndrome related to a human disease
CC      associated with the polypeptides e.g. cancer. ABUS2311-ABUS2408 represent
XX      the human NOVX proteins of the invention
XX
SQ      Sequence 525 AA;
XX

Alignment Scores:
Pred. No.:      7,97e-90      Length:      525
Score:          1191.00      Matches:      222
Percent Similarity: 99.55%      Conservative: 0
Best Local Similarity: 99.55%      Mismatches: 0
Query Match:    28.16%      Indels:      1
DB:             6      Gaps:      0

US-10-028-952a-4 (1-2333) x ABUS2367 (1-525)
QY      1095 ACCACCTCCATCTCCAGTCTCCAGCCCTGAGAGAGAGCCCTTCTGACAGACAG 1154
Db      304 ThrThrsertileserserProGlnProGlyLysLeuArgserProPheLeuGlnYgIn 323
QY      1155 CTCACCCACCAAGAGAGACCACTTTGGCAGAGAGCCAGCTGTCGCATCTCAAGGCCAGG 1214
Db      324 LeuthrgInProGluThrHisPheGlyArgGluProAlaAlaIleSerArgProArg 343
QY      1215 GCAGATCTCCCTGCTGAGAGAGCGGCGCCAGACCTCCTCATGTCTGTGAGGACAGAA 1274
Db      344 AlaaspLeuProAlaGluGluProAlaProSerThrProCysLeuValGlnAlaGlu 363
QY      1275 GAGGAGCGCTGTGTATGAGAACTCCAGACAGAGACCTTCTACAGACAGCCCCCACTG 1334
Db      364 GluGluAlaValTyrGluGluProProGluGluGlnThrPheTyrGluGlnProProLeu 383
QY      1335 GTGCAGACAGAGAGTGTGCTGCTGACAGATTCAGACACACATTCAGGGCCAGGAGGCTC 1394
Db      384 ValGlnGlnGlnGlyAlaGlySerGlnHisIlePheHisIleGlnGlnGlnGlyLeu 403
QY      1395 AGTGGAGAGAGGCTCTGTGCCGCTGCGCTGTACACTTACAGAGAGCCAGACAGACAGAG 1454
Db      404 SerGlyGlnGlyLeuGlyAlaArgAlaLeuTyrAspTyrGlnAlaAlaAspAspThrGlu 423
QY      1455 ATCTCCTTTGACCCCGGAAACCTCATCAGGGGATCGAGGATTCAGACAGAGGCTGATG 1514
Db      424 IleSerPheAspProGluAsnLeuIleThrGlyIleGluValIlePheGlnGlyTyrPTr 443
QY      1515 COTGCTATGGGCGGAGATGGCCATTTTGGCATGTTCCCTCCCACTACGTGAGACTCATT 1574
```

Db 444 ArgGlyTyrGlyProAspGlyHisPhe-AlaCysSerLeuProThrThrTrpSerSerLe 463  
QY 1575 GAGTGAAGGCTGAGGAGCAATCTTCCCTCCCTTCACACATGCTTCTTATGCTGG 1634  
Db 463 uSerGluAlaGluGlyThrSerCysProSerProLeuArgHisGlyPheLeuIleAlaG 483  
QY 1635 AAGAGAGGCGCTGAGATTGACATTCAGCACTCTTCAGAGAAATAGAGCCCGAGTGAAGA 1694  
Db 483 yArgGlyGlyLeuGlyValAspIleGlnHisSerSerArgAsnArgThrProSerGlnAs 503  
QY 1695 TGAAGCCTCAGGCGCTCCCTCCGCTTGGCAGATTCAGCTGTACCCCAATGCGAGCAT 1754  
Db 503 pGluAlaSerGlyLeuProProAlaTrpGlnThrGlnProValThrProAsnAlaIaIa 523  
QY 1755 GGCCTGG 1761  
Db 523 tAlaTrp 525

RESULT 4  
AAW05398  
ID AAW05398 standard; protein; 355 AA.  
XX AAW05398;  
XX 19-FEB-1998 (first entry)  
XX Human clone 56 protein.  
XX Src-homology region 3 domain; human; mouse; SH3 domain; cell growth;  
XX cellular signalling element; cellular structural element; malignancy;  
XX protein identification; functional domain; protein screening;  
XX cellular signal transduction process.  
XX Homo sapiens.  
XX MO9631625-A1.  
XX 10-OCT-1996.  
XX 04-APR-1996; 96WC-US004454.  
XX 07-APR-1995; 95US-00417872.  
XX 03-APR-1996; 96US-00630915.  
XX (CYTO-) CYTOGEN CORP.  
XX (UTNC-) UNIV NORTH CAROLINA.  
XX Sparks AB, Hoffman N, Kay BK, Fowlkes DM, McConnell SJ;  
XX WPI: 1996-465045/46.  
XX N-PSDB: AAT39798.  
XX Identifying polypeptide(s) having specific functional domain (esp. SH3  
XX domain) - comprises detecting selective binding to recognition unit,  
XX regardless of sequence homology.  
XX Claim 54; Fig 57; 174pp; English.

CC AAW05386-W05403 represent novel human and mouse Src-homology region 3  
CC (SH3) domain containing proteins that can be used in the method of the  
CC invention. SH3 domain containing proteins play a role in signalling and  
CC structural elements of cells. The method of the invention is for  
CC identifying polypeptides containing functional domains of interest  
CC (especially SH3 domains). The method comprises contacting a multivalent  
CC recognition unit (RU) complex with a number of peptides and identifying  
CC polypeptides having a selective binding affinity for the RU complex. The  
CC method is based on functional similarities and does not rely on sequence  
CC similarities. Prior methods only gave limited success for identifying  
CC proteins which contain an SH3 domain due to the minimal sequence homology  
CC among known SH3 proteins. It has been found that small peptide RUs in  
CC multivalent form have reduced specificity for a given functional domain  
CC compared to monomer RUs. Multivalent RU complexes are particularly suited  
CC to screening for polypeptides containing functional domains that are

CC similar to, but not identical in sequence to, the original target  
CC functional domain. The new method enables proteins having a common  
CC function to be identified. Identification of novel SH3 proteins will be  
CC useful for a better understanding of cell growth, malignancy, signal  
CC transduction processes, etc. New candidate drugs can be identified, and  
CC their specificities (e.g. pharmacological activities) can be assessed  
CC using the method of the invention

XX SQ Sequence 355 AA;  
XX

Alignment Scores:  
Pred. No.: 9,51e-90 Length: 355  
Score: 1189.50 Matches: 251  
Percent Similarity: 73.28% Conservative: 15  
Best Local Similarity: 69.15% Mismatches: 63  
Query Match: 28.12% Indels: 34  
Gaps: 6  
DB: 2

US-10-028-952A-4 (1-233) x AAW05398 (1-355)

QY 724 CTCAGTACGAGAGAACCCGACGTTTCAGACATTGCTGATG-----TGCTTG 771  
Db 12 LeuProGlyArgGlyThrProGlyProSerGlyLeuCysValProGluAspGlnCysArg 31  
QY 772 GCTGAGGAGCAATGGGGCGAAGCTACCATCTGTGGGATATGACTGAACGCTCTAAGT 831  
Db 32 ValArgAspLeuLysGlyThrLeuAspSerPheThrAlaLysAlaGluLysGluGlu 51  
QY 832 CAGATCCCGCCGACGAGCGAAAGATACGAGCGCCCGGAGCCTCGTGGCTCGGAT 891  
Db 52 AsnArgArgLeuGluGluLysArgTrpAlaGluLysAlaGlnArgGlnLeuGluGlu 71  
QY 892 AGCGGTCGCCCGCTGTCGCCCGGCGGCGCGCCGCCCTCAGCGCGCCGCGCGC 951  
Db 72 ArgArgGluArgGluLeu-----ArgGluAlaAlaArgArgGluGlnArgTyrGlnGlu 89  
QY 952 GCGGAGGAGC--GCGTGCCTCCCGCGCGCGC----- 979  
Db 90 GlnGly-GlyGluAlaSerProGlnSerArgThrTrpGluGlnGlnGluValAlaIse 109  
QY 980 -CGGACCGGAGTCCGCGCGGAGTGCCTTCGTCGGAAGAGCGGCGCGGAGAA 1038  
Db 109 rArgAsnArgAsnGluGlnGluSerAlaValHisProArgGluIle-PheLysGlnLysG 129  
QY 1039 GCGGCGCGCCCGCTGCGCCGTCAGCAACGACGTTGCTGGGAAACCTGGCGTAAACCA 1098  
Db 129 LuArg-----AlaMetSer-----ThrT 135  
QY 1099 CCTCATCTCCAGATCTCTTACGCTTGAGAGAGTGAAGAGCCCTTCTTGAGAGAGAGCTCA 1158  
Db 135 hrSerIleSerSerProGlnProGlyLysLeuArgSerProPheLeuLysGlnLeu 155  
QY 1159 CCCAGCCAGAGCCGACCTTGGCAGAGCGAGCTGCGCATCTGATCTGAGGCGAG 1218  
Db 155 hGlnProGluThrHisPheGlyArgGluProAlaAlaAlaIleSerArgProArgAla 175  
QY 1219 ATCTCCCTGCTGAGAGCGCGCGCCGACGACTCTTCATGCTGTGTGAGGCGAAGAG 1278  
Db 175 splLeuProAlaGluLubProAlaProSerTrpProPheCysLeuValGlnAlaGluGlu 195  
QY 1279 AGCGTGTGATGAGAGACCTCCAGAGCGAGAGACCTTTCAGAGAGAGCGCCGACTGTGC 1338  
Db 195 LuAlaValTyrGluLubProPheGluGlnGluThrPheTyrGlnGlnProPheLeuValG 215  
QY 1339 AGCAGCAAGTCTGCTGTGAGCATTTGACCACTTGAAGGCGAGGCGGCTCACTG 1398  
Db 215 InGlnGlnGlyAlaGlySerGlnHisIleAspHisHisIleGlnGlnGlnLysSerG 235  
QY 1399 GGCAGAGGCTGTGTCGCGTGCCTGTACGACTTACAGGAGCGAGCAGCAGAGATCT 1458  
Db 235 LysGlnGlyLeuLysAlaArgAlaLeuTyrAspTyrGlnAlaAlaAspAspThrGlnLys 255  
QY 1459 CCTTGAACCCGAGAACTCATACGAGGATCGAGAGTATGACAGAAAGGCTGTGGCTG 1518

```

Db      255 ePheASPProGluAsnLeuIleThrGlyIleGluValIleAspGluGlyTyrTrpArgG 275
QY      1519 GCTATGGCGCCGATGGCCATTTCGATGTTCCCTCCCACTACGTGAGCTCATGAGT 1578
Db      275 IYTYGlyProAspGlyHisPheGlyMetPheProAlaAsnTyrValGluLeuIleAsp- 294
QY      1579 GAGGCTGAGGAGCACTTCGCTCCCTCCCTCAGACAGCTGCTTCCTATGCTGAGAGA 1638
Db      295 GluAlaGluGlyThrSerCysProSerProLeuA-GHLSGlyPheLeuIleHISGlyArg 314
QY      1638 GAGAGCTGAGATTGACATTCAGCACTCTCCAGAAATAGACCCCGCATGAGAGTAG 1698
Db      315 GlyGlyLeuGlyValAlaSpIleGlnHisSerSerArgAsnArgThrProSerGluAspGlu 334
QY      1699 GCCTCAGGCGCTCCCTCCGCTTGCGAGACTCAGCTGTGACCCCAATGCGCAATGGCC 1758
Db      335 AlaSerGlyLeuProProAlaTyrGlnThrGlnProValThrProAlaAlaValGlu 354
QY      1759 TGG 1761
Db      355 Trp 355

```

## RESULT 5

AA23756 standard; protein; 538 AA.

AA23756;

09-SEP-1999 (first entry)

Human mammary epithelial cell growth inhibitor.

Human mammary epithelial cell growth inhibitor.

Human; mammary; breast cancer.

Homo sapiens.

W0932625-A2.

01-JUL-1999.

18-DEC-1998; 98WO-US027147.

19-DEC-1997; 97US-00994076.

(UNMT) UNIV MICHIGAN.

Ervin PR;

WPI; 1998-418924/35.

N-PSDB; AA85833.

DNA encoding Mammary epithelial cell growth inhibitor.

useful for therapy of breast cancer.

The present sequence represents human mammary epithelial cell growth inhibitor.

CC by normal human mammary cells in culture and recombinant expressed

Mammary epithelial cell growth inhibitor, and is

useful as a therapeutic agent in the treatment of breast cancer

Sequence 538 AA;

Alignment Scores:

Pred. No.: 1.03e-76

Score: 1033.00

Percent Similarity: 95.52%

Best Local Similarity: 95.07%

Query Match: 24.42%

DB: 2

US-10-028-952a-4 (1-2333) X AA23756 (1-538)

```

QY      417 TTGATCTTGATTTTTCAGTACGAATACAGCCGGAAGGCGGCGCTCAGATCCTCTGA 476
Db      1 MetIleuIlePheSerThrAsnTrpAspArgGluSerGlyAlaSer-Asp-LeuLeu 20
QY      477 CCTTTGGGTTTTCAGAGAGGCTGTCAAGAAAGTTTACACAGAGATTAAGCTGCTGTGG 536
Db      20 hrHerPValIleuSerArgGlyHisSerThrHisArgAspAsnTrpLeuValA 40
QY      537 CGGCGAAGCTTTCATACGACCGCTGCTTTTTCATCCTTGATGCTGCTGCTTTCAT 596
Db      40 IalAlaYsArgSerLysArgArgPheLeuIleuArgCysArgLeuPheLeuSerL 60
QY      597 TGTGAAGCAGATTCACCAAGCGTGGATTTTCACCACTAATAGGAACGTGAGCTGG 656
Db      60 euGlySerArgIleHisGlnAlaLeuAspCysSerProThrAsnArgGluArgGluLeuG 80
QY      657 GATTAGACCGCTCGTGAAGAGCTTATGTTTACCTCTGATGATGATGATGATGATGAT 716
Db      80 ILeuAspArgArgGluThrGly-LeuPheThrLeuLeuMetCysCysGlyHisGly 99
QY      717 AATCTGCTCAGTACGAGAGAGAACCGCAGGTTCCAGATTTGCTGATGCTGCTGCTGA 776
Db      100 AsnProAlaGlnTyrGluArgAsnArgArgPheArgHisLeuValTyrValLeuGlyTyr 119
QY      777 GAGCCCAATGGGCGGAAGCTTACCATCTGTGGATTTGACTGAACGCTTACGAGAA 836
Db      120 GlyAlaAsnGlyAlaLeuLeuProSerValGlyLeuLeuLeuAsnAlaSerLysSerGlu 139
QY      837 TCCCGCCAGAGCGGAAGATACGACGCGCGGAGGCTCGGTTGAGCTCGGATAGCG 896
Db      140 SerArgProGlyGlyThrIleArgGlnArgArgGlyAlaSerValGlyLeuGlyTyrPro 159
QY      897 GTCCCGCGCTGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 955
Db      160 ValProArgLeuSerProProAlaGlyAlaProProProProArgAlaProAlaArg 179
QY      956 GAGGCGCGTGGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1015
Db      180 GluGlyAlaCysProAlaAlaArgArgAspArgGlyProValArgSerAlaLeuArgPro 199
QY      1016 GGGAAACGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 1075
Db      200 GlyLysArgGlyAlaAlaGlyLysAlaAlaAlaProSerProValThrHisArgThrPhe 219
QY      1076 GTG 1078
Db      220 Val 220

```

## RESULT 6

AB804724 standard; protein; 302 AA.

AB804724;

11-MAR-2002 (first entry)

Human PPS423 protein SEQ ID NO:26.

Human; PPS423; cancer suppression.

Homo sapiens.

CN131316-A.

19-SEP-2001.

13-MAR-2000; 2000CN-00111990.

13-MAR-2000; 2000CN-00111990.

(SHAN-) SHANGHAI INST ONCOLOGY.

XX

PI Gu J, Yang S;  
 XX WPI, 2002-042194/06.  
 DR N-PSDB; ABA04459.  
 XX  
 PT New human protein able to suppress growth of cancer cells and its  
 PT encoding polynucleotide.  
 XX  
 PS Claim 1; Page 35 (Disclosure); 36pp; Chinese.  
 CC The present sequence represents human PP5423 protein, which has cancer  
 CC suppressing activity. The present invention describes a method for the  
 CC preparation of the protein by recombination, and the application of the  
 CC protein in treating diseases such as cancer  
 XX  
 XX Sequence 302 AA;

Alignment Scores:  
 Pred. No.: 1,98e-63 Length: 302  
 Score: 872.00 Matches: 179  
 Percent Similarity: 74.21% Conservative: 8  
 Best Local Similarity: 71.03% Mismatches: 57  
 Query Match: 20.61% Indels: 9  
 DB: 5 Gaps: 2

US-10-028-952a-4 (1-2333) X ABB04724 (1-302)

QY 841 GCCACGCGGACGATACGCGAGCGCGGAGCTCGGTTGGCTCGGATAGCCGCTCC 900  
 DB 53 AAlLylhlaagluys-----GlugluGluaaAargluegluGlulysAargla 70  
 QY 901 CCCGCTGTCCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGGAGG 960  
 DB 71 GlugluAaGluaAargluAargluAargluAargluAargluAargluAargluAa 90  
 QY 961 CGGTCGGAA 1020  
 DB 91 AargluegluAargluAargluAargluAargluAargluAargluAargluAarglu 110  
 QY 1021 ACGGG-----GCGCGCGCGGAAAGCGCGCGCGCGCGCGCGCGCGCGCGTCA 1061  
 DB 111 GlugluGluaValaValaAarglaAarglaAarglaAarglaAarglaAargla 130  
 QY 1062 CGGACCGCGAGCTCGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGT 1121  
 DB 131 GluilepheylsGluysGluAarglaAarglaAarglaAarglaAarglaAargla 150  
 QY 1122 GCGAAGCTGAGAGCCCTTCTGCGAAGAGCGCGCGCGCGCGCGCGCGCGCGCGT 1181  
 DB 151 GllyleuAargserProPheleuGluysGluAarglaAarglaAarglaAargla 170  
 QY 1182 AGAGAGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1241  
 DB 171 AargluegluAarglaAarglaAarglaAarglaAarglaAarglaAarglaAargla 190  
 QY 1242 CCAGAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1301  
 DB 191 ProSerThrProProcySleuValaGluAargluAargluAargluAargluAarglu 210  
 QY 1302 GAGCAGAGAGCTTTCTAAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1361  
 DB 211 GlugluGluaAargluAargluAargluAargluAargluAargluAargluAarglu 230  
 QY 1362 CACATGACACCACTTCAAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1421  
 DB 231 HistleAspHisHisHisHisHisHisHisHisHisHisHisHisHisHisHis 250  
 QY 1422 CTGTAGACTTACAGAGAGCGCGAGAGAGAGATCTCTTGAACCCCGAAGCTCAGT 1481  
 DB 251 LeuTyrAspTyrGlnAlaAarglaAarglaAarglaAarglaAarglaAargla 270  
 QY 1482 ACGGGATGAGGTGATCGAGAGGTGCTGCGCGCGCGCGCGCGCGCGCGCGCGT 1541

DB 271 ThGlyleGluaValleAspGluGlyTrpTrpAargluGlyTyrGluProAspGluPhe 290  
 QY 1542 GGCATGTTCCCTGCCCACTACGTGAGCTCATTTAG 1577  
 DB 291 GlyMetPheProAlaAsnTyrValGluLeuileGlu 302

RESULT 7  
 AAM78585  
 ID AAM78585 standard; protein: 430 AA.

AC AAM78585;  
 XX 06-NOV-2001 (first entry)  
 DT  
 XX  
 XX Human protein SEQ ID NO 1247.

KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KM tissue growth factor; immunomodulatory; cancer; leukaemia;  
 OS nervous system disorder; arthritis; inflammation.

OS Homo sapiens.  
 RN WO200157190-A2.  
 XX  
 XX 09-AUG-2001.  
 PD  
 XX 05-FEB-2001; 2001WO-US004098.

PR 03-FEB-2000; 2000US-00496914.  
 PR 27-APR-2000; 2000US-00560875.  
 PR 20-JUN-2000; 2000US-00598075.  
 PR 19-JUL-2000; 2000US-00620325.  
 PR 01-SEP-2000; 2000US-00654936.  
 PR 15-SEP-2000; 2000US-00663561.  
 PR 20-OCT-2000; 2000US-00693325.  
 PR 30-NOV-2000; 2000US-00728422.

PA (HISE-) HISEQ INC.

PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;  
 PI Ma Y, Zhao Qa, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;

DR WPI, 2001-476283/51.  
 DR N-PSDB; AAK51718.

PT Nucleic acids encoding polypeptides with cytokine-like activities, useful  
 PT in diagnosis and gene therapy.

PS Claim 20; Page 3507-3508; 6221pp; English.

CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activity/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the  
 CC sequence listing were missing at the time of publication

SO Sequence 430 AA;

Alignment Scores:  
 Pred. No.: 2.18e-63 Length: 430  
 Score: 872.00 Matches: 179  
 Percent Similarity: 74.21% Conservative: 8  
 Best Local Similarity: 71.03% Mismatches: 57



Query Match: 20.61% Indels: 9  
 DB: 4 Gaps: 2  
 US-10-028-952a-4 (1-2333) x AAM78585 (1-430)

QY 841 GCCCAGCGGAGAACATACGCGAGCGCCGCGAGCCTCGGTTGGCCTCGATAGCCGGTCC 900  
 DB 181 AAlaYsAlaGluYs-----GluGluGluYsnaYsnaYsGluGluYsnaYsnaYsAla 198  
 QY 901 CCCGCTGTCCCGCGCGCGCGCGCCCGCTCCACGCGCGCGCGCGCGCGCGGAGG 960  
 DB 199 GluGluAlaGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 218  
 QY 961 CGCGTGGCCCGGAA 1020  
 DB 219 ArgArgGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 238  
 QY 1021 ACGGG-----CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTCA 1061  
 DB 239 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 258  
 QY 1062 CGCAGCG 1121  
 DB 259 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 278  
 QY 1122 GCGAAGCTGAGAGCG 1181  
 DB 279 GlyGlySerArgSerProGluGluGluGluGluGluGluGluGluGluGluGluGlu 298  
 QY 1182 AGAGAGCG 1241  
 DB 299 ArgGluProAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 318  
 QY 1242 CCAGAGCTCCCG 1301  
 DB 319 ProSerThrProProCysLeuValGluGluGluGluGluGluGluGluGluGluGlu 338  
 QY 1302 GAGCAGAGACCTTCTTACAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1361  
 DB 339 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 358  
 QY 1362 CACATGACACACACATTCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1421  
 DB 359 HisIleSerHisHisIleGluGluGluGluGluGluGluGluGluGluGluGluGlu 378  
 QY 1422 CTGTACGACTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1481  
 DB 379 LeuTyAspTyrGlnAlaAlaAlaAspThrGluIleSerPheAspProGluAsnLeu 398  
 QY 1482 ACCGCGCATCGAGTGTATCCAGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1541  
 DB 399 ThrGlyIleGluValIleAspGluGluGluGluGluGluGluGluGluGluGlu 418  
 QY 1542 GGCATGTTCCCTGCACTACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1577  
 DB 419 GlyMetPheProAlaAsnTyrValGluLeuIleGlu 430

RESULT 8  
 AAM79569 ID AAM79569 standard; protein; 458 AA.  
 XX AAM79569;  
 XX  
 DT 06-NOV-2001 (first entry)  
 XX Human protein SEQ ID NO 3215.  
 XX  
 KM Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
 KM vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
 KM tissue growth factor; immunomodulatory; cancer; leukaemia;  
 KM nervous system disorder; arthritis; inflammation.  
 XX  
 OS Homo sapiens.

XX  
 PN W0200157190-A2.  
 XX  
 PD 09-AUG-2001.  
 XX  
 PF 05-FEB-2001; 2001WC-US004098.  
 XX  
 PR 03-FEB-2000; 2000US-00496914.  
 PR 27-APR-2000; 2000US-00560875.  
 PR 20-JUN-2000; 2000US-00598075.  
 PR 19-JUL-2000; 2000US-00620325.  
 PR 01-SEP-2000; 2000US-00654936.  
 PR 15-SEP-2000; 2000US-00663561.  
 PR 20-OCT-2000; 2000US-00693325.  
 PR 30-NOV-2000; 2000US-00728422.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;  
 PI Ma Y, Zhao QA, Wang D, Wang U, Zhang J, Ren F, Chen R, Wang ZW;  
 PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;  
 XX  
 DR WPI: 2001-476283/51.  
 DR N-PSDB; AAK52702.  
 XX  
 PT Nucleic acids encoding polypeptides with cytokine-like activities, useful  
 in diagnosis and gene therapy.  
 XX  
 PS Claim 20; Page 287; 6221pp; English.  
 XX  
 CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
 CC encoded polypeptides (AAM78323-AAK80302) that exhibit activity elating to  
 CC cytokine, cell proliferation or cell differentiation or which may induce  
 CC production of other cytokines in other cell populations. The  
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
 CC peptide therapy. The polypeptides have various cytokine-like activities,  
 CC e.g. stem cell growth factor activity, haematopoiesis regulating  
 CC activity, tissue growth factor activity, immunomodulatory activity and  
 CC activity/inhibin activity and may be useful in the diagnosis and/or  
 CC treatment of cancer, leukaemia, nervous system disorders and/or  
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
 CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the  
 CC sequence listing were missing at the time of publication  
 XX  
 SQ Sequence 458 AA;  
 XX

Alignment Scores:  
 Pred. No.: 2.22e-63 Length: 458  
 Score: 872.00 Matches: 179  
 Percent Similarity: 74.21% Conservative: 8  
 Best Local Similarity: 71.03% Mismatches: 57  
 Query Match: 20.61% Indels: 9  
 DB: 4 Gaps: 2

US-10-028-952a-4 (1-2333) x AAM79569 (1-458)

QY 841 GCCCAGCGGAGAACATACGCGAGCGCCGAGCCTCGTGGCTCGATAGCCGGTCC 900  
 DB 209 AAlaYsAlaGluYs-----GluGluGluYsnaYsnaYsGluGluYsnaYsnaYsAla 226  
 QY 901 CCCGCTGTCCCGCGCGCGCGCGCCCGCTCCACGCGCGCGCGCGCGCGCGGAGG 960  
 DB 227 GluGluAlaGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 246  
 QY 961 CGCGTGGCCCGGAA 1020  
 DB 247 ArgArgGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 266  
 QY 1021 ACGGG-----CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTCA 1061  
 DB 267 GluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGluGlu 286  
 QY 1062 CGCAGCG 1121

Db	287	GluIlePheIysGInLysGlnAAlaWaseSerThrTrsSerIleSerSerProGlnPro	106
QY	1122	GGCAAGCTGAGGAGACCCCTTCCTGCAGAGCAAGCTCAACCAACAGAGACCCTTTGGC	1181
Db	307	GlyLysLeuArgSerProPheLeuGInLysGInLeuThrGInProGInThrHisPheGly	326
QY	1182	AGAGAGCCAGCTGCTGGCCATCTCAAGGCCACGGGCGAGATCTCCCTGCCTGAGAGACCGGCG	1241
Db	327	ArgGInProAlaAlaAlaAlaIleSerArgProArgAlaAspLeuProAlaGlnGInProAla	346
QY	1242	CCCAAGCACTCCCATGTCTGTGTGAGGACAGAAAGAGAGGCTGTATAGAGAACTTCCA	1301
Db	347	ProSerThrProProCysLeuValGlnIlaGlnGInGInAlaValItyrGlnGlnProPro	366
QY	1302	GAGAGAGAGACCTTCTTACAGACAGGCCCCACCTGGGCGACAGCAAGGCTGGGCTCTGAG	1361
Db	367	GlnGlnGlnThrPheItyrGlnGlnProProLeuValGlnGlnGlnGlnItyrAlaGlySerGln	386
QY	1362	CACATTGACCAACCACTTCAGGGCCAGGGGGCTCACTGAGGCGCAAGGCTCTGTGCCCTGCC	1421
Db	387	HisIleAspHisIleHisIleGlnGlnGlnItyrLeuSerGlyGlnGlyLeuCysAlaArgAla	406
QY	1422	CTGTACGACTACCGAGGACGCGACAGACAGAGATCTCTTGTGACCCCGAGAACTCATC	1481
Db	407	LeuItyrAspItyrGlnAlaAlaIleAspAspThrGlnIleSerPheAspProGlnAsnLeuIle	426
QY	1482	ACGGGCGATCGAGGTGATGCAGCAAGGCTGTGGCCTGGCTATGGCCGATGGCCATTTT	1541
Db	427	ThrGlyIleGlnValIleAspGlnGlyItyrTrpArgGlyItyrGlyProAspGlyHisPhe	446
QY	1542	GGCATGTCTCCCGCAACTACGCTGAGGCTCATTTGAG	1577
Db	447	GlyMetCpHeProAlaAsnItyrValGlnLeuIleGln	458
RESULT 9			
ID	AAB63214		
XX	AAB63214	standard; protein; 199 AA.	
XX	AAB63214;		
XX	26-MAR-2001	(first entry)	
DE		Human secreted protein sequence encoded by gene 38 SEQ ID NO:140.	
XX		Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;	
KW		antirheumatic; antiproliferative; cyclostatic; cardiant; vasotropic;	
KW		antibiochemic; nootropic; neuroprotective; antibacterial; virucide;	
KW		fungicide; ophthalmologically; gene therapy; autoimmune disease; neoplasm;	
KW		rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;	
KW		cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;	
KW		angiogenesis; nervous system disorder; Alzheimer's disease; infection;	
KW		ocular disorder; corneal infection; wound healing; skin aging;	
XX		food additive; preservative.	
XX	OS	Homo sapiens.	
XX	PM	WO200061628-A1.	
XX	PD	19-OCT-2000.	
XX	PF	06-APR-2000; 2000WO-US009071.	
XX	PR	09-APR-1999; 99US-0128694P.	
XX	PR	20-JAN-2000; 2000US-0176931P.	
PA		(HUMA-) HUMAN GENOME SCI INC.	
PA		(ROSE/) ROSEN C A.	
XX		Ruben SM, Komatsoulis G;	
XX	PI		
XX	DR	WPI; 2000-647420/62.	

FT		Isolated nucleic acid molecule encoding a human secreted protein is used
FT		in preventing, treating or ameliorating a medical condition.
XX		
XX		Disclosure; Page 518-519; 533pp; English.
PS		
CC	AAP22373 to AAP22421 encode the human secreted proteins given in AAB61334	
CC	to AAB61382. AAB61383 to AAB613231 represent more human secreted proteins	
CC	and polypeptides homologous to them. Human secreted proteins have	
CC	activities based on the tissues and cells the genes are expressed in.	
CC	Examples of activities include: immunosuppressive; antiarthritic;	
CC	antirheumatic; antiproliferative; cytostatic; cardiant; vasotropic;	
CC	cerebroprotective; neurotropic; neuroprotective; antibacterial; virucide;	
CC	fungicide; and ophthalmological. The polynucleotides and proteins can be	
CC	used to prevent, treat or ameliorate a medical condition in e.g. humans,	
CC	mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are	
CC	also used in diagnosing a pathological condition or susceptibility to a	
CC	pathological condition. Disorders which are diagnosed or treated include	
CC	autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative	
CC	disorders e.g. neoplasms of the breast or liver, cardiovascular disorders	
CC	e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,	
CC	angiogenesis, nervous system disorders e.g. Alzheimer's disease,	
CC	infections caused by bacteria, viruses and fungi and ocular disorders	
CC	e.g. corneal infection. The polypeptides can also be used to aid wound	
CC	healing and epithelial cell proliferation, to prevent skin aging due to	
CC	sunburn, to maintain organs before transplantation, for supporting cell	
CC	culture of primary tissues, to regenerate tissues and in chemotaxis. The	
CC	polypeptides can also be used as a food additive or preservative to	
CC	increase or decrease storage capabilities. AAP22364 to AAP22372 and	
CC	AAB61333 represent sequences used in the exemplification of the present	
CC	invention	
XX		
SQ	Sequence 199 AA:	
	Alignment Scores:	
Pred. No.:	2.58e-63 Length: 199	
Score:	870.00 Matches: 161	
Percent Similarity:	100.00% Conservative: 0	
Best Local Similarity:	100.00% Mismatches: 0	
Query Match:	20.57% Indels: 0	
DB:	Gaps: 0	
US-10-028-952A-4 (1-2333) x AAB63214 (1-199)		
OY	1095 ACCACCTTCATCTCCAGTCTCAGCTCGGACGAGGTGAAGGCCCTTCCTGCAGAACAG 1154	
Db	39 ThrThrSerIleSerSerProGln.ProGlyLeuArgSerProPheLeuGlnLyGln 58	
OY	1155 CTCACCCCAACAGAGACCATTGGTCAGAGAAGCCAGCTGCTGCATCTCAAAGGCCAG 1214	
Db	59 LeuthrGrnPogIutnRhIsphagLIARgInuProAlaAlaAlIeserAsProArg 78	
OY	1215 GCAGATCCCCCTGGTGAGAGACCGGCGCCACGACTCCTCCATGTCGTGTGACAGCAAA 1274	
Db	79 AlaAspleuProAlaGInuInuProAlaProserThrProProCySleuValGlnAlaGlu 98	
OY	1275 GAGAGAGCTGTGTATGAGAACTTCACAGAGAGACCTTCACGAGCACCCCCTAGT 1334	
Db	99 GluGlnAlaValTYrGluGInuPrProGluGInuGlnuTrPheTYrGluGInuProleu 118	
OY	1335 GTGCAGCAGCAAGGTGTGCTGTGAGACCATTCACACACATTCAGAGGCCAGGGGCTC 1394	
Db	119 ValGInuGInuGInuYAlaGlySeoCIuhItIleAspHisIleGInuLYGlInuLYeu 138	
OY	1395 AGTGGGCAAGGCTTGTGCCCCCGCTGTACACTACACGAGCGGACGAGCAAGAG 1454	
Db	139 SerGIyGInuLYeuCYAlaArgLIaleutyryspRYrGlnAlaAlaAspAspHrnGlu 158	
OY	1455 ATCTCCTTTGACCCCGAGAACCTTCACGCGGACTCAGAGGTGATGACGAAGGCTGTGG 1514	
Db	159 IleSerPheAspProGluAsnLeuIleThrgIyIleGluValIleAspGluGLYrTpTrp 178	
OY	1515 CGTGGCTATGGGCGGAGTGGCATTTGGCATGTCCTCGCAATTAAGTGGAGCTGATT 1574	





Db 257 AlaAspLeuProAlaGluGluProAlaProSerThrProProCysLeuValGlnAlaGlu 276  
 QY 1275 GAGAGAGCTGTGTATGAGAACTCCAGACGAGAGACCTTCTTACAGACAGCCCCCACTG 1334  
 Db 277 GluGlnAlaValTyrGluGluProProGluGlnGluThrPheTyrGluGlnProProLeu 296  
 QY 1335 GTGACAGCAGAGGTGTGCTGTGACCATTTGACCACTTACAGGAGGCGAGGCTC 1394  
 Db 297 ValGlnGlnGlnGlnAlaGlySerGlnHisIleAspHisIleGlnGlnGlnGlu 316  
 QY 1395 AGTGGGCAAGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 1454  
 Db 317 SerGlyGlnGlyLeuGlnCysAlaArgAlaLeuTyrAspTyrGlnAlaAlaAspThrGlu 336  
 QY 1455 ATCTCTTTCACCCCGAGACCTCATCAGCGGAGTGTGAGTGTGAGGAGGCTGTG 1514  
 Db 337 IleSerPheAspProGluAsnLeuIleThrGlyIleGlnValIleAspGlnGlyTyrTrp 356  
 QY 1515 CGTGGCTATGAGCGCGATGCGCATTTTGCGATGTTCCTGCGCACTACGTGAGCTCAT 1574  
 Db 357 ArgGlyTyrGlyProAspGlyHisPheGlyMetPheProAlaAsnTyrValGluLeuIle 376  
 QY 1575 GAG 1577  
 Db 377 Glu 377

## RESULT 13

AA820896  
 ID AAB20896 standard; protein; 431 AA.

XX AAB20896;

XX 08-JAN-2001 (first entry)

XX Human dreblin-like protein and SH3 domain sequence SEQ ID NO:1.

XX Human, dreblin-like protein; SH3 domain; cancer; diagnosis.

XX Homo sapiens.

XX JP2000197489-A.

XX 18-JUL-2000.

XX 07-JAN-1999; 99JP-00002254.

XX 07-JAN-1999; 99JP-00002254.

XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.

XX WPI; 2000-545912/50.

XX N-PSDB; AAA92163; AAA92164.

XX New human protein having dreblin-like sequence and SH3 domain.

XX Claim 1; Page 5-6; 12p; Japanese.

XX The present sequence represents a human protein having a dreblin-like

XX sequence and SH3 domain. The protein and antibody against the protein

XX can be used in the diagnosis and the treatment of various diseases

XX including cancers

XX SQ Sequence 431 AA;

## Alignment Scores:

Pred. No.: 3.2e-63 Length: 431  
 Score: 870.00 Matches: 161  
 Percent Similarity: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 20.57% Indels: 0  
 DB: 3 Gaps: 0

US-10-028-952a-4 (1-233) x AAB20896 (1-431)

QY 1095 ACCACCTCCATCTCCAGTCTCCAGCTGTGAGAGCTGAGAGCCCTTCTCTGACAGACAG 1154  
 Db 271 ThrThrSerIleSerSerProGlnProGlyLysLeuArgSerProPheGlnGlnGln 290  
 QY 1155 CTCACCCACACAGAGACCCCACTTTTGCAAGAGAGGCACTGTGCAATCTCAAGCCCAAG 1214  
 Db 291 LeuThrGlnProGluThrHisPheGlyArgGluProAlaAlaIleSerArgProArg 310  
 QY 1215 GCAATCTCTCTGTGAGAGACCGGCGCCGACCACTCTCATGTCTGTGAGAGAGAA 1274  
 Db 311 AlaAspLeuProAlaGluGluProAlaProSerThrProProCysLeuValGlnAlaGlu 330  
 QY 1275 GAGAGAGCTGTGTATGAGAACTCCAGACGAGAGACCTTCTTACAGACAGCCCCCACTG 1334  
 Db 331 GluGlnAlaValTyrGluGluProProGluGlnGluThrPheTyrGluGlnProProLeu 350  
 QY 1335 GTGACAGCAGCAGGTGCTGTGCTGTGACCATTTGACCACTTACAGGAGGCGAGGCTC 1394  
 Db 351 ValGlnGlnGlnGlnAlaGlySerGlnHisIleAspHisIleGlnGlnGlnGlu 370  
 QY 1395 AGTGGGCAAGGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTG 1454  
 Db 371 SerGlyGlnGlyLeuGlnCysAlaArgAlaLeuTyrAspTyrGlnAlaAlaAspThrGlu 390  
 QY 1455 ATCTCTTTCACCCCGAGACCTCATCAGCGGAGTGTGAGTGTGAGGAGGCTGTG 1514  
 Db 391 IleSerPheAspProGluAsnLeuIleThrGlyIleGlnValIleAspGlnGlyTyrTrp 410  
 QY 1515 CGTGGCTATGAGCGCGATGCGCATTTTGCGATGTTCCTGCGCACTACGTGAGCTCAT 1574  
 Db 411 ArgGlyTyrGlyProAspGlyHisPheGlyMetPheProAlaAsnTyrValGluLeuIle 430  
 QY 1575 GAG 1577  
 Db 431 Glu 431

## RESULT 14

AA85562  
 ID AA85562 standard; protein; 431 AA.

XX AA85562;

XX 12-FEB-2001 (first entry)

XX Human tyrosine kinase substrate cKs118/Dresh protein sequence.

XX Tyrosine kinase substrate; Tks 107, GRUB, Tks 113, Tks 118, Dresh,

XX Tks 202; Src; rheumatoid arthritis; atherosclerosis; stroke; cancer;

XX autoimmune disorder; organ transplantation; myocardial infarction;

XX cardiomyopathy; renal failure; neurodegenerative disorder;

XX cardiovascular disorder; inflammatory bowel disease; multiple sclerosis;

XX asthma; osteoarthritis; psoriasis; rhinitis.

XX Homo sapiens.

XX WO200061750-A2.

XX 19-OCT-2000.

XX 06-APR-2000; 2000WO-US009277.

XX 09-APR-1999; 99US-0128492P.

XX (SUGC-) SUGEN INC.

XX Phan H, Courtneldge SA;

XX WPI; 2000-679486/66.

XX N-PSDB; AAC66078.

PT Novel tyrosine kinase substrate polypeptides and polynucleotides, used to

diagnose and treat diseases such as rheumatoid arthritis,

PT artherosclerosis and cancer.

PS Claim 2; Fig 7; 120pp; English.

This invention relates to isolated, enriched or purified nucleic acid molecules represented by sequences AAG6076-C66079 which encode tyrosine kinase substrate (Tks) Tks 107/GRB, Tks 113, Tks 118/Dresh or Tks 202 represented by sequences AY85650-Y85661. The Tks proteins are substrates for cytoplasmic tyrosine kinase Src. The invention encompasses probe sequences, recombinant cells, antibodies specific for the Tks proteins, and hybridomas producing the antibodies. The Tks proteins exhibit antihepatic; antiallergic; antiarteriosclerotic; immunosuppressive; cardiatic; cerebroprotective; neuroprotective; cytosratic; antiinflammatory; osteoparatic; antiporatic; antiallergic, and antasthmatic activities. The tyrosine kinase substrates are used to treat and diagnose diseases or disorders selected from Rheumatoid arthritis, arteriosclerosis, autoimmune disorders, organ transplantation, myocardial infarction, cardiomyopathies, stroke, renal failure, oxidative cancer-related neurodegenerative disorders, cardiovascular disorders, cancer, and immune-related disorders selected from chronic inflammatory bowel disease, chronic inflammatory pelvic disease, multiple sclerosis, asthma, osteoarthritis, psoriasis, and rheinitis. The proteins can also be used to identify modulators of its activity. DNA sequences encoding the proteins may be used in gene therapy methods, and to produce transgenic animals.

Sequence 431 AA;

Alignment Scores:

Pred. No.:	3.2e-3	length:	411
Score:	870.00	Matches:	161
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	20.57%	Indels:	0
DB:	3	Gaps:	0

US-10-028-952A-4 (1-2333) x AAY85662 (1-431)

QY	1095	ACCACCTGCATATCCAGACCTCCACCCCTGGACCGAAGAGTACGAGAGCCCTTCTGTGAGAAAGAG	1154
Db	271	TTTTTSeTLeSeSeSeProGInProGILyLeuNuSeSeProPheNeuNuInySeIn	290
QY	1155	CTCACCCAAACAGAGAGCCCACTTTGGAGAGAGCCAGCTGCTGCATCTCAAGCCCAAG	1214
Db	291	LeuTrTgInProGILuThThIaPheGILyArGILuProLaLaLaLaLeSeTrArProArG	310
QY	1215	GGAGATCTCCCTGCTGAGAGAGCCGAGCCGACCACTCTCATGTCTGTGAGAGCAADA	1274
Db	311	AlaAePLeuProLaILGILuILuProLaIaProSeTrHProProCyLeuValGILuLeGILu	330
QY	1275	GAGGAGGCGCTGTATAGAGAAACCTCCAGAGAGAGACCTTCAAGAGAGAGGCCCACTG	1334
Db	331	GIUGILuAaValTyGILuGILuProProGILuGILuThrPheTyxILuInProTroleu	350
QY	1335	GTGACAGACGAAAGTGCTGCTCTGAGACATTGACCACTTCAGGGCCAGGGGCTC	1394
Db	351	ValGILuGILuGILuValAGILySeGILuThIaIaPheSILuThIaIaLeGILuInILyInILyLeu	370
QY	1395	AGTGGGCAAGGGCTTGTGCTCCGTCGCTGTACGACTACAGGACACCGACACAGACAGAG	1454
Db	371	SeGILyGILuGILyLeuCyAlaIaArgIaLeuNuTyzSPILyGILuIaIaIaAaPheSPHrGILu	390
QY	1455	ATCTCCTTTGAGACCCGAGAACTCATACACGGGCACTCAGAGTGATGTCACAGAAAGCTGTGG	1514
Db	391	ILeSePheAePProGILuAaSeuLeuILeThrCylILeILuValILeAaPILuILyTrHPr	410
QY	1515	CGTGGCTATAGGGCCGGATGGCCATTGGCATGTTCTCTGCCAACTACGTGAGACTATT	1574
Db	411	ArGGLyTyGILyProAaPILyHILaPheILyMeChPheTrolaAaSeTyValGILuLeuILe	430
QY	1575	GAG 1577	
Db	431	GIu 431	





GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2004, 10:23:34 ; Search time 53.8053 Seconds

(without alignments)  
4164.275 Million cell updates/sec

Title: US-10-028-952a-9

Perfect score: 3989

Sequence: 1 HEIPVPTVTPAKPQXKXERA.....DFKSKKKKKLKLMTSSX 793

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : 1: Genesecp29Jn04:\*  
2: genesecp1980s:\*  
3: genesecp2000s:\*  
4: genesecp2001s:\*  
5: genesecp2002s:\*  
6: genesecp2003as:\*  
7: genesecp2003bs:\*  
8: genesecp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3535	88.6	760	4 AAB66394	Aab66394 Human pro
2	1193.5	29.9	753	4 AAB66395	Aab66395 Human mam
3	800.5	20.1	355	2 AAM05398	Aam05398 Human CLO
4	796	20.0	199	3 AAB63214	Aab63214 Human sec
5	796	20.0	203	5 ABB57455	Abb57455 Human sec
6	796	20.0	217	2 AAY73941	Aay73941 Human pro
7	796	20.0	302	5 ABB04724	Abb04724 Human PPS
8	796	20.0	430	4 AAM78585	Aam78585 Human pro
9	796	20.0	431	3 AAB20896	Aab20896 Human dre
10	796	20.0	431	3 AAY85662	Aay85662 Human tyr
11	796	20.0	458	4 AAM79569	Aam79569 Human pro
12	794	19.9	377	6 ABB98850	Abb98850 Human str
13	794	19.9	439	4 AAB93895	Aab93895 Human str
14	746.5	18.7	525	6 ABB52367	Abb52367 Human GPC
15	647	16.2	162	2 AAY73942	Aay73942 Human pro
16	628.5	15.8	433	2 AAM05389	Aam05389 Mouse SH3
17	626.5	15.7	201	3 AAB63213	Aab63213 Gene 38 h
18	409	10.3	538	2 AAY23756	Aay23756 Human mam
19	300	7.5	58	2 AAM34223	Aam34223 SH3 doma
20	300	7.5	58	2 AAM34248	Aam34248 SH3 doma
21	241	6.0	55	3 AAB53794	Aab53794 Human cor
22	239.5	6.0	301	4 AAM83949	Aam83949 Human imm
23	234	5.9	486	6 ABR44110	Abr44110 Human pro
24	234	5.9	486	6 ABO07214	Abo07214 Human p53
25	234	5.9	486	7 ADD67624	Add67624 Human Ly1

26	231	5.8	559	4 ABB60041	Abb60041 Drosophil
27	228.5	5.7	586	2 AAY03795	Aay03795 Murine lc
28	227.5	5.7	431	4 ABB71831	Abb71831 Drosophil
29	221.5	5.6	486	2 AAM05408	Aam05408 Mouse HSI
30	209	5.2	471	7 ADE55094	Ade55094 Rat Prote
31	208	5.2	546	6 AAM05406	Aam05406 Mouse cor
32	208	5.2	546	6 AAE38430	Aae38430 Mouse cor
33	205	5.1	456	6 AAE38429	Aae38429 Human cor
34	205	5.1	513	7 ADE55096	Ade55096 Human Pro
35	202	5.1	550	6 ABO07213	Abo07213 Human p53
36	192	4.8	40	3 AAB53737	Aab53737 Human col
37	180.5	4.5	53	2 AAY03828	Aay03828 LckBP1 al
38	177.5	4.4	53	6 AAE38444	Aae38444 Human cor
39	177.5	4.4	53	6 AAE38445	Aae38445 Mouse cor
40	176.5	4.4	54	2 AAY03827	Aay03827 LckBP1 al
41	175.5	4.4	54	2 AAY03829	Aay03829 LckBP1 al
42	171.5	4.3	53	2 AAY03821	Aay03821 LckBP1 al
43	171.5	4.3	53	2 AAY03826	Aay03826 LckBP1 al
44	171.5	4.3	53	2 AAY03820	Aay03820 LckBP1 SH
45	170.5	4.3	53	2 AAY03824	Aay03824 LckBP1 al

## ALIGNMENTS

RESULT 1  
AAB66394 standard; protein; 760 AA.  
XX XX  
XX AAB66394;  
XX XX  
DT 10-APR-2001 (first entry)  
XX XX  
DE Human prostate ECGI protein sequence.  
XX XX  
KW Human; mammaplastin A; mammaplastin B; mammaplastin C; ECGI; cancer;  
XX XX  
KW epithelial cell growth inhibitor.  
XX XX  
OS Homo sapiens.  
XX XX  
PN WO200078955-A1.  
XX XX  
PD 28-DEC-2000.  
XX XX  
PF 19-JUN-2000; 2000WO-US016900.  
XX XX  
PR 16-JUN-1999; 99US-0139995P.  
XX XX  
PA (BIOT-) BIOTHERAPIES INC.  
XX XX  
PI Ervin PR;  
XX XX  
DR WPI, 2001-071393/08.  
XX XX  
DR N-PSDB; AAF31282.  
XX XX  
PT A Mammaplastin-like epithelial cell growth inhibitor, and the nucleic that  
PT encodes it, useful for diagnosing and/or preventing epithelial cell  
PT cancers, e.g. of the ovaries or prostate.  
XX XX  
PS Disclosure; Page 41-43; 55pp; English.  
XX XX  
CC The present invention describes a mammaplastin-like epithelial cell growth  
CC inhibitor (ECGI) which has substantial similarity to mammaplastin A, B or  
CC C. ECGI is expressed in healthy cells, but is either absent or expressed  
CC at reduced levels in cancerous cells. The protein and its coding sequence  
CC can be used to inhibit epithelial cell growth and the amount present in  
CC cells can be used to diagnose cancer or monitor its progress  
XX XX  
SQ Sequence 760 AA;  
XX XX  
Query Match 88.6%; Score 3535; DB 4; Length 760;  
Best Local Similarity 95.8%; Pred. No. 0;  
Matches 759; Conservative 0; Mismatches 1; Indels 32; Gaps 30;

QY 1 HEIPIVPIYPAKPOKXERAMNQGKTLSTLWHEEETXYXNKKYAPQAPVSPRG 60  
 DB 1 HEIPIVPIYPAKPOKXERAMNQGKTLSTLWHEEETXYXNKKYAPQAPVSPRG 60  
 QY 61 ARGGXRPCCGPVKYHYSDRFTDPRVRCGEPRGALASXAPRPAARRPGATRGXYYRWG 120  
 DB 61 ARGGXRPCCGPVKYHYSDRFTDPRVRCGEPRGALASXAPRPAARRPGATRGXYYRWG 120  
 QY 61 ARGGXRPCCGPVKYHYSDRFTDPRVRCGEPRGALASXAPRPAARRPGATRGXYYRWG 120  
 DB 61 ARGGXRPCCGPVKYHYSDRFTDPRVRCGEPRGALASXAPRPAARRPGATRGXYYRWG 120  
 QY 121 VXLGRTYCTVXVXSGELEDNPMKXAPAKLILFSTNTDXXSAGSEYFPGYKA 180  
 DB 121 VXLGRTYCTVXVXSGELEDNPMKXAPAKLILFSTNTDXXSAGSEYFPGYKA 180  
 QY 181 GAVKATYTGITGMRSVSDVAFXSPDVSSYHXXAEFTKRWIVHPLIGXSMXDVYVR 240  
 DB 176 G-VKATYTGITGMRSVSDVAFXSPDVSSYHXXAEFTKRWIVHPLIGXSMXDVYVR 232  
 QY 241 QVSFTLLMKCCCHGNPAQYERNRXXHLYTVLGXANAKLXSVGLXNASKSRRPGXT 300  
 DB 233 QVSFTLLMKCCCHGNPAQYERNRXXHLYTVLGXANAKLXSVGLXNASKSRRPGXT 290  
 QY 301 ISORRASVGLGXEXXRLSPAPRPPHAPRGRARAPRAGTGVGCXVFPVYLGNGAR 360  
 DB 291 ISORRASVGLGXEXXRLSPAPRPPHAPRGRARAPRAGTGVGCXVFPVYLGNGAR 348  
 QY 361 ERFPKXRPRTARSWGTWRTYXISISPOPKLRSFLOXOLTOPTHGREDPAAXRPR 420  
 DB 349 ERFPKXRPRTARSWGTWRTYXISISPOPKLRSFLOXOLTOPTHGREDPAAXRPR 404  
 QY 421 ADLPAEPAPSPPCIVQAEBAVYEEPEXCEFTPEQCPPLVQOQXGSEHIDH1QOQGL 480  
 DB 405 ADLPAEPAPSPPCIVQAEBAVYEEPEXCEFTPEQCPPLVQOQXGSEHIDH1QOQGL 462  
 QY 481 SNQGLCARLIDYQADDEISFDRENILITGIEVXXSAGWAGYGPDDHFGKXPAHYELI 540  
 DB 463 S-QGICARLIDYQADDEISFDRENILITGIEVXXSAGWAGYGPDDHFGKXPAHYELI 519  
 QY 541 EXGKGHILAFPSQWLPYCMKRRPXXXSALFOEXDPOKXGLRAPSGLADSAHPKCN 600  
 DB 520 EXGKGHILAFPSQWLPYCMKRRPXXXSALFOEXDPOKXGLRAPSGLADSAHPKCN 577  
 QY 601 GLVTPHPCIPRXXXTAMLPITGYXAKPCWBPSPPLPSCEGSGXGASGLMTPSA 660  
 DB 578 GLVTPHPCIPRXXXTAMLPITGYXAKPCWBPSPPLPSCEGSGXGASGLMTPSA 634  
 QY 661 FICLXXFSLASGWWPPLFRXNLGNSERRELFLAEFTVYXRVVDHGLAAGNLSCXHLCA 720  
 DB 635 FICLXXFSLASGWWPPLFRXNLGNSERRELFLAEFTVYXRVVDHGLAAGNLSCXHLCA 691  
 QY 721 PHSISLGLGXGKXGCRWPSHPRYSKADTTGSGTFLITCLQXVYCASTDSDFRSGK 780  
 DB 692 PHSISLGLGXGKXGCRWPSHPRYSKADTTGSGTFLITCLQXVYCASTDSDFRSGK 749  
 QY 781 KXKLEKLTSS 792  
 DB 750 K-KLEKLTSS 760

RESULT 2  
 AAB66395  
 ID AAB66395 standard; protein; 753 AA.  
 AC AAB66395;  
 DT 10-APR-2001 (first entry)  
 DE Human mammaplastin C protein sequence.  
 KW Human; mammaplastin A; mammaplastin B; mammaplastin C; ECGI; cancer;  
 KW epithelial cell growth inhibitor.  
 OS Homo sapiens.  
 XX

PN W0200078955-A1.  
 XX 28-DEC-2000.  
 PD 19-JUN-2000; 2000WO-US016900.  
 PF 18-JUN-1999; 99US-0139995P.  
 PR (BIOT-) BIOTHERAPIES INC.  
 PA Evin PR;  
 PI WPI, 2001-071393/08.  
 DR A Mammaplastin-like epithelial cell growth inhibitor, and the nucleic that  
 PT encodes it, useful for diagnosing and/or preventing epithelial cell  
 PT cancers, e.g. of the ovaries or prostate.  
 PS Disclosure, Page 44-46; 55pp; English.  
 XX The present invention describes a mammaplastin-like epithelial cell growth  
 CC inhibitor (ECGI) which has substantial similarity to mammaplastin A, B or  
 CC C. ECGI is expressed in healthy cells, but is either absent or expressed  
 CC at reduced levels in cancerous cells. The protein and its coding sequence  
 CC can be used to inhibit epithelial cell growth and the amount present in  
 CC cells can be used to diagnose cancer or monitor its progress  
 XX Sequence 753 AA;  
 SQ  
 Query Match 29.9%; Score 1193.5; DB 4; Length 753;  
 Best Local Similarity 81.6%; Pred. No. 1.6e-113;  
 Matches 280; Conservative 1; Mismatches 29; Indels 33; Gaps 13;  
 QY 37 HGEETXZYXNWXKAPGAPPVSPRAGGXXRPGCPVKYHYSDRFTDPRRCGERGAL 96  
 DB 5 HGEETXZYXNWXKAPGAPPVSPRAGGXXRPGCPVKYHYSDRFTDPRRCGERGAL 62  
 QY 97 ASXAKRPAARPGATRGSGXARWGVXGRTYCTVXVXSGELEDNPMKXAPAKLIL 156  
 DB 63 AG-AKRPAAARPGATRGSGXARWGVXGRTYCTVXVXSGELEDNPMKXAPAKLIL 119  
 QY 157 LIFSTNTDXXSAGARSTYPPGFYXGKXRTTGTGTGMRSVSDVAFXSPDVSSYHXX 216  
 DB 120 LIFSTNTDXXSAGARSTYPPGFYXGKXRTTGTGTGMRSVSDVAFXSPDVSSYHXX 177  
 QY 217 AEFTKRWIVHPLIGXSMXDMXTVVRQVSFTLLMKCCCHGNPAQYERNRXXHLYTVL 276  
 DB 178 AEFTKRWIVHPLIGXSMXDMXTVVRQVSFTLLMKCCCHGNPAQYERNRXXHLYTVL 235  
 QY 277 NGAKLXSVGLXNASKSRRPGKTIIRORRASVGLGXPPXRLSPAPRPPHAPRGRAR 336  
 DB 236 NGAKLXSVGLXNASKSRRPGKTIIRORRASVGLGXPPXRLSPAPRPPHAPRGRAR 290  
 QY 337 RAPPAGTGVRCXVPVYLGNGARPERRPXXRPRTARSWGTWR 379  
 DB 291 -----GGRVPRAPGPGSACPSRWETGR 314

RESULT 3  
 AAM05398  
 ID AAM05398 standard; protein; 355 AA.  
 AC AAM05398;  
 DT 19-FEB-1998 (first entry)  
 DE Human clone 56 protein.  
 KW Src-homology region 3 domain; human; mouse; SH3 domain; cell growth;  
 KW cellular signaling element; cellular structural element; malignancy;  
 KW protein identification; functional domain; protein screening;  
 KW cellular signal transduction process.  
 XX

OS Homo sapiens.  
 XX MO9631625-A1.  
 PN 10-OCT-1996.  
 PD 10-OCT-1996.  
 PF 04-APR-1996; 96WO-US004454.  
 XX 07-APR-1995; 95US-00417872.  
 PR 03-APR-1996; 96US-00630915.  
 XX (CYTO-) CYTOGEN CORP.  
 PA (UNIC-) UNIV NORTH CAROLINA.  
 XX Sparks AB, Hoffman N, Kay BK, Fowlkes DW, McConnell SJ;  
 PI WPI; 1996-465045/46.  
 DR N-PSDB; AAT39798.  
 XX Identifying polypeptide(s) having specific functional domain (esp. SH3  
 PT domain) - comprises detecting selective binding to recognition unit,  
 PT regardless of sequence homology.  
 XX Claim 54; Fig 57; 174p; English.  
 XX AAW05386-W05403 represent novel human and mouse Src-homology region 3  
 CC (SH3) domain-containing proteins that can be used in the method of the  
 CC invention. SH3 domain-containing proteins play a role in signalling and  
 CC structural elements of cells. The method of the invention is for  
 CC identifying polypeptides containing functional domains of interest  
 CC (especially SH3 domains). The method comprises contacting a multivalent  
 CC recognition unit (RU) complex with a number of peptides and identifying  
 CC polypeptides having a selective binding affinity for the RU complex. The  
 CC method is based on functional similarities and does not rely on sequence  
 CC similarities. Prior methods only gave limited success for identifying  
 CC proteins which contain an SH3 domain due to the minimal sequence homology  
 CC among known SH3 proteins. It has been found that small peptide RUs in  
 CC multivalent form have reduced specificity for a given functional domain  
 CC compared to monomer RUs. Multivalent RU complexes are particularly suited  
 CC to screening for polypeptides containing functional domains that are  
 CC similar to, but not identical in sequence to, the original target  
 CC functional domain. The new method enables proteins having a common  
 CC function to be identified. Identification of novel SH3 proteins will be  
 CC useful for a better understanding of cell growth, malignancy, signal  
 CC transduction processes, etc. New candidate drugs can be identified, and  
 CC their specificities (e.g. pharmacological activities) can be assessed  
 CC using the method of the invention  
 CC  
 CC Sequence 355 AA;  
 SQ  
 Query Match 20.1%; Score 800.5; DB 2; Length 355;  
 Best Local Similarity 64.9%; Pred. No. 1.9e-73;  
 Matches 157; Conservative 9; Mismatches 47; Indels 29; Gaps 1;  
 QY 359 RPERPXXRPRTRASNGTWTXSTISSPGKRSRPTLOXLOPETHREPREAAXSR 418  
 DB 112 RNEQSAVHPREIFQKERAMSTSTISPOGKRSRPLQOLQOPETHREPREAAXSR 171  
 QY 419 PRADLPAREPAPSPXPCLVQAEAEVVEEPEYEOETFEYEQPPLVQOXXGSHIHDIHQG 478  
 DB 172 PRADLPAREPAPSPXPCLVQAEAEVVEEPEYEQPPLVQOXXGSHIHDIHQG 231  
 QY 479 GLSXGGLARALYDQADDXEISFDPENLTGTBYXXEGWWRGYPDGHGKMPAYVE 538  
 DB 232 GLSXGGLARALYDQADDXEISFDPENLTGTBYXXEGWWRGYPDGHGKMPAYVE 291  
 QY 539 LIEXXGHIILXP-----SQTWLPYCKRRAPXXXS 569  
 DB 292 LIIDEAGTSCPSPLRHGFLINGRGLGYDIGHSSNNRTPRSDEASGLPPAWQTQPTV PNA 351  
 QY 570 AL 571  
 DB 352 AM 353

RESULT 4  
 AAB63214  
 ID AAB63214 standard; protein; 199 AA.  
 XX  
 XX AAB63214;  
 AC 26-MAR-2001 (first entry)  
 XX  
 DT Human secreted protein sequence encoded by gene 38 SEQ ID NO:140.  
 DE  
 XX Human; secreted protein; diagnosis; immunosuppressive; antiarthritic;  
 KW antirheumatic; antiproliferative; cytosstatic; cardiant; vasotropic;  
 KW cerebroprotective; noctropic; neuroprotective; antibacterial; virucide;  
 KW fungicide; opthalmological; gene therapy; autoimmune disease; neoplasm;  
 KW rheumatoid arthritis; hyperproliferative disorder; cardiac arrest;  
 KW cardiovascular disorder; cerebrovascular disorder; cerebral ischaemia;  
 KW angiogenesis; nervous system disorder; Alzheimer's disease; infection;  
 KW ocular disorder; corneal infection; wound healing; skin aging;  
 KW food additive; preservative.  
 XX  
 OS Homo sapiens.  
 XX  
 XX WO200061623-A1.  
 XX 19-OCT-2000.  
 XX 06-APR-2000; 2000WO-US009071.  
 XX 09-APR-1999; 95US-0128694P.  
 XX 20-JAN-2000; 2000US-0176931P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA (ROSE/) ROSEN C A.  
 PA Ruben SM, Komatsoulis G;  
 PI WPI; 2000-647420/62.  
 DR Isolated nucleic acid molecule encoding a human secreted protein is used  
 PT in preventing, treating or ameliorating a medical condition.  
 PT  
 XX Disclosure; Page 518-519; 533p; English.  
 XX AAF22373 to AAF22421 encode the human secreted proteins given in AAB63134  
 CC to AAB63182. AAB63183 to AAB63231 represent more human secreted proteins  
 CC and polypeptides homologous to them. Human secreted proteins have  
 CC activities based on the tissues and cells the genes are expressed in.  
 CC Examples of activities include: immunosuppressive; antiarthritic;  
 CC antirheumatic; antiproliferative; cytosstatic; cardiant; vasotropic;  
 CC cerebroprotective; noctropic; neuroprotective; antibacterial; virucide;  
 CC fungicide; and opthalmological. The polynucleotides and proteins can be  
 CC used to prevent, treat or ameliorate a medical condition in e.g. humans,  
 CC mice, rabbits, goats, horses, cats, dogs, chickens or sheep. They are  
 CC also used in diagnosing a pathological condition or susceptibility to a  
 CC pathological condition. Disorders which are diagnosed or treated include  
 CC autoimmune diseases e.g. rheumatoid arthritis, hyperproliferative  
 CC disorders e.g. neoplasms of the breast or liver, cardiovascular disorders  
 CC e.g. cardiac arrest, cerebrovascular disorders e.g. cerebral ischaemia,  
 CC angiogenesis, nervous system disorders e.g. Alzheimer's disease,  
 CC infections caused by bacteria, viruses and fungi and ocular disorders  
 CC e.g. corneal infection. The polypeptides can also be used to aid wound  
 CC healing and epithelial cell proliferation, to prevent skin aging due to  
 CC sunburn, to maintain organs before transplantation, for supporting cell  
 CC culture of primary tissues, to regenerate tissues and in chemotaxis. The  
 CC polypeptides can also be used as a food additive or preservative to  
 CC increase or decrease storage capabilities. AAF22364 to AAF22372 and  
 CC AAB63133 represent sequences used in the exemplification of the present  
 CC invention  
 XX  
 XX Sequence 199 AA;  
 SQ

Query Match 20.0%; Score 796; DB 3; Length 199;  
 Best Local Similarity 82.5%; Pred. No. 2.6e-73;  
 Matches 151; Conservative 2; Mismatches 30; Indels 0; Gaps 0;

QY 359 RPERFXKRPRTARSGTWRTXSSISPPQKLSPPLOXQLOTPETHFGREPAAXXSR 418  
 DB 17 RNEQSAVHPRRIFFQKRAMSTTSSISPPQKLSPPLOXQLOTPETHFGREPAAISR 76  
 QY 419 PRADLPABEPAPSPPCLVQAEAEVYEEPEQETFFEQPPLVQOQXGSEHIDHIOGQ 478  
 DB 77 PRADLPABEPAPSPPCLVQAEAEVYEEPEQETFFEQPPLVQOQXGSEHIDHIOGQ 136  
 QY 479 GLSXQGLCARALYDYOADDXEISPPENLITGIEVXXEGMWRGYPDGHFGMPANYVE 538  
 DB 137 GLSGGGLCARALYDYOADDXEISPPENLITGIEVDEGWMWRGYPDGHFGMPANYVE 196  
 QY 539 LIE 541  
 DB 197 LIE 199

## RESULT 5

AB57455  
 ID AB57455 standard; protein; 203 AA.

AC AB57455;  
 XX

DT 15-MAR-2002 (first entry)  
 XX

DE Human secreted protein encoding polypeptide SEQ ID NO 101.  
 XX

KW Human; noctropic; neuroprotective; cytostatic; dermatological; vinctide;  
 KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnery;  
 KW antiparkinsonian; antischizoid; antianaemic; antichitric; cancer;  
 KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;  
 KW antiallergic; antidiabetic; antilucer; anticonvulsant; antifungal;  
 KW antiparastic; cardiant; immune disorder; cardiovascular disorder;  
 KW neurological disease; infection; nephrotropic; gene therapy; vaccine.

OS Homo sapiens.  
 XX

PN MO200183510-A1.  
 XX

PD 08-NOV-2001.  
 XX

PF 26-APR-2001; 2001WO-US013318.  
 XX

PR 02-MAY-2000; 2000US-0201194P.  
 XX

PR 16-JUN-2000; 2000US-0212142P.  
 XX

PA (HUMA-) HUMAN GENOME SCI INC.  
 XX

PI Komatsoulis G, Ruben SM, Rosen CA;  
 XX

DR WPI; 2002-121886/16.  
 XX

PT An isolated nucleic acid molecule encoding a human secreted protein (SP)  
 XX

PT Alzheimer's disease and cancers.  
 XX

PS Disclosure; Page 491; 496pp; English.  
 XX

XX The invention relates to novel genes (AB101566-AB101594) and proteins  
 CC (AB57394-AB57456) useful for preventing, treating or ameliorating  
 CC medical conditions e.g. by protein or gene therapy. The genes are  
 CC isolated from a range of human tissues disclosed in the specification.  
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful in  
 CC the diagnosis, treatment and prevention of: (a) cancer, e.g. breast and  
 CC ovarian cancer and other cancers of the adrenal gland, bone, bone marrow,  
 CC breast, gastrointestinal tract, liver, lung, or urogenital; (b) immune  
 CC disorders e.g. Addison's disease, allergies, autoimmune haemolytic  
 CC anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's disease,  
 CC multiple sclerosis, rheumatoid arthritis and ulcerative colitis; (c)

CC cardiovascular disorders such as myocardial ischaemia; (d) wound healing  
 CC (e) neurological diseases e.g. cerebral anoxia and epilepsy; and (f)  
 CC infectious diseases such as viral, bacterial, fungal and parasitic  
 CC infections

XX Sequence 203 AA;  
 XX

QY 359 RPERFXKRPRTARSGTWRTXSSISPPQKLSPPLOXQLOTPETHFGREPAAXXSR 418  
 DB 21 RNEQSAVHPRRIFFQKRAMSTTSSISPPQKLSPPLOXQLOTPETHFGREPAAISR 80  
 QY 419 PRADLPABEPAPSPPCLVQAEAEVYEEPEQETFFEQPPLVQOQXGSEHIDHIOGQ 478  
 DB 81 PRADLPABEPAPSPPCLVQAEAEVYEEPEQETFFEQPPLVQOQXGSEHIDHIOGQ 140  
 QY 479 GLSXQGLCARALYDYOADDXEISPPENLITGIEVXXEGMWRGYPDGHFGMPANYVE 538  
 DB 141 GLSGGGLCARALYDYOADDXEISPPENLITGIEVDEGWMWRGYPDGHFGMPANYVE 200  
 QY 539 LIE 541  
 DB 201 LIE 203

## RESULT 6

AA73941  
 ID AA73941 standard; protein; 217 AA.

XX AA73941;  
 AC

DT 14-MAR-2000 (first entry)  
 XX

DE Human prostate tumor EST fragment derived protein #128.  
 XX

KW Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;  
 KW treatment.  
 KW

XX Homo sapiens.  
 XX

PN DE19820190-A1.  
 XX

PD 04-NOV-1999.  
 XX

PF 28-APR-1998; 98DE-01020190.  
 XX

PR 28-APR-1998; 98DE-01020190.  
 XX

PA (META-) METAGEN GES GENOMFORSCHUNG MBH.  
 XX

PI Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E;  
 XX

DR WPI; 1999-621386/54.  
 XX

DR N-PSDB; AA52900.  
 XX

PT New human nucleic acid sequences from pancreatic tumors, and related  
 XX

PT proteins.  
 XX

PS Claim 23; Page 362; 502pp; German.  
 XX

XX This invention describes novel polypeptides and their encoding nucleic  
 CC acids derived from human pancreatic tumor tissue which have cytostatic  
 CC activity. The sequences are also useful in producing pharmaceutical  
 CC compositions for treatment of pancreatic tumors. AA73814-774252  
 CC represent protein fragments encoded by the human pancreatic tumor cDNA  
 CC library derived expressed sequence tag (EST) sequences represented in  
 CC AA52858-253014

XX Sequence 217 AA;  
 SQ

Query Match 20.0%; Score 796; DB 2; Length 217;

Best Local Similarity 82.5%; Pred. No. 2.9e-73; Matches 151; Conservative 2; Mismatches 30; Indels 0; Gaps 0;

QY 359 RPERPPKRRPSRTKRSCTGTXYSISPPQKLRSPFLQXLTQPTHTGRRPAXXR 418  
DB 35 RNEQSAVHPRRIKOKERAMSTISISSPQKLRSPFLQXLTQPTHTGRRPAXISR 94  
QY 419 PRADLPAPBPAPSPXPCIVQAEAEVAYEPEEOTFEYQOPPLVQOQXGSEHIDHIOGQ 478  
DB 95 PRADLPAPBPAPSPXPCIVQAEAEVAYEPEEOTFEYQOPPLVQOQXGSEHIDHIOGQ 154  
QY 479 GLSXQGLCARALYDYQAADDXEISFPDENLITGIEVXXEGMWRGYPGDFGKXPANYVE 538  
DB 155 GLSGQGLCARALYDYQAADDTETISFPDENLITGIEVIDEGMWRGYPGDFGMPANYVE 214  
QY 539 LIE 541  
DB 215 LIE 217

## RESULT 7

ABB04724 ID ABB04724 standard; protein; 302 AA.

XX AC ABB04724;  
XX DT 11-MAR-2002 (first entry)  
XX DD Human PPS423 protein SEQ ID NO:26.  
XX KW Human; PPS423; cancer suppression.  
XX OS Homo sapiens.  
XX PN CN131316-A.  
XX PD 19-SEP-2001.  
XX PR 13-MAR-2000; 2000CN-00111990.  
XX PR 13-MAR-2000; 2000CN-00111990.  
XX PR 13-MAR-2000; 2000CN-00111990.  
XX PA (SHAN-) SHANGHAI INST ONCOLOGY.  
XX PI Gu J, Yang S;  
XX DR WPI; 2002-042194/06.  
XX DR N-PSDB; ABA04459.  
XX PT New human protein able to suppress growth of cancer cells and its  
XX PT encoding polynucleotide.  
XX PS Claim 1; Page 35 (Disclosure); 38pp; Chinese.  
XX CC The present sequence represents human PPS423 protein, which has cancer  
XX CC suppressing activity. The present invention describes a method for the  
XX CC preparation of the protein by recombinant, and the application of the  
XX CC protein in treating diseases such as cancer  
SQ Sequence 302 AA;

Query Match 20.0%; Score 796; DB 5; Length 302;

Best Local Similarity 82.5%; Pred. No. 4.5e-73; Matches 151; Conservative 2; Mismatches 30; Indels 0; Gaps 0;

QY 359 RPERPPKRRPSRTKRSCTGTXYSISPPQKLRSPFLQXLTQPTHTGRRPAXXR 418  
DB 120 RNEQSAVHPRRIKOKERAMSTISISSPQKLRSPFLQXLTQPTHTGRRPAXISR 179  
QY 419 PRADLPAPBPAPSPXPCIVQAEAEVAYEPEEOTFEYQOPPLVQOQXGSEHIDHIOGQ 478  
DB 180 PRADLPAPBPAPSPXPCIVQAEAEVAYEPEEOTFEYQOPPLVQOQXGSEHIDHIOGQ 239

QY 479 GLSXQGLCARALYDYQAADDXEISFPDENLITGIEVXXEGMWRGYPGDFGKXPANYVE 538  
DB 240 GLSGQGLCARALYDYQAADDTETISFPDENLITGIEVIDEGMWRGYPGDFGMPANYVE 299

QY 539 LIE 541  
DB 300 LIE 302

## RESULT 8

AAM78585 ID AAM78585 standard; protein; 430 AA.

XX AC AAM78585;  
XX DT 06-NOV-2001 (first entry)  
XX DD Human protein SEQ ID NO 1247.  
XX KW Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
XX KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
XX KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
XX KW nervous system disorder; arthritis; inflammation.  
XX OS Homo sapiens.  
XX PN WO200157190-A2.  
XX PD 09-AUG-2001.  
XX PF 05-FEB-2001; 2001WO-US004098.  
XX PR 03-FEB-2000; 2000US-00486914.  
XX PR 27-APR-2000; 2000US-00560875.  
XX PR 20-JUN-2000; 2000US-00598075.  
XX PR 19-JUL-2000; 2000US-00620325.  
XX PR 15-SEP-2000; 2000US-00654936.  
XX PR 15-SEP-2000; 2000US-00653561.  
XX PR 20-OCT-2000; 2000US-00693325.  
XX PR 30-NOV-2000; 2000US-00728422.  
XX PA (HYSB-) HYSEQ INC.  
XX PI Tang YF, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;  
XX PI Ma Y, Zhao Q, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;  
XX PI Xue AJ, Yang Y, Wehrman T, Goodrich R;  
XX DR WPI; 2001-476283/51.  
XX DR N-PSDB; AAK51718.  
XX PT Nucleic acids encoding polypeptides with cytokine-like activities, useful  
XX PT in diagnosis and gene therapy.  
XX PS Claim 20; Page 3507-3508; 622pp; English.  
XX CC The invention relates to polynucleotides (AAK51456-AAK53435) and the  
XX CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to  
XX CC cytokine, cell proliferation or cell differentiation or which may induce  
XX CC production of other cytokines in other cell populations. The  
XX CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
XX CC peptide therapy. The polypeptides have various cytokine-like activities,  
XX CC e.g. stem cell growth factor activity, haematopoiesis regulating  
XX CC activity, tissue growth factor activity, immunomodulatory activity and  
XX CC activity/inhibin activity and may be useful in the diagnosis and/or  
XX CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
XX CC inflammation. Notes: Records for SEQ ID NO 2110 (AAK52581), 2111  
XX CC (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the  
XX CC sequence listing were missing at the time of publication

SQ Sequence 430 AA;

Query Match 20.0%; Score 796; DB 4; Length 430;

Best Local Similarity 82.5%; Pred. No. 7.1e-73; Indels 0; Gaps 0;  
Matches 151; Conservative 2; Mismatches 30;

QY 359 RPERRPXRRPRTARSMGTWRTXSSISPPGKLRSPFLQXLTQTPETHFGREPAAXSR 418  
| | | | |  
DB 248 RNEQESAVHPREIFKQKERAMSTSSISPPGKLRSPFLQXLTQTPETHFGREPAAXSR 307  
| | | | |

QY 419 PRADLPAAEPAPSPPCLVQAEBAVYEEPEQSTTFEOPPLVQOQXGSEHIDHIIQOQ 478  
| | | | |  
DB 308 PRADLPAAEPAPSPPCLVQAEBAVYEEPEQSTTFEOPPLVQOQXGSEHIDHIIQOQ 367  
| | | | |

QY 479 GLSXQGLCARALYDQAAADXEISFDPENLITGIEVXXEGMWRGYPGDGHFGKXPANYVE 538  
| | | | |  
DB 368 GLSXQGLCARALYDQAAADXEISFDPENLITGIEVXXEGMWRGYPGDGHFGKXPANYVE 427  
| | | | |

QY 539 LIE 541  
| | | | |  
DB 428 LIE 430  
| | | | |

RESULT 9  
AAB20896  
ID AAB20896 standard; protein; 431 AA.  
XX  
AC AAB20896;  
XX  
DT 08-JAN-2001 (first entry)  
XX  
DE Human dreblin-like protein and SH3 domain sequence SEQ ID NO:1.  
XX  
KW Human; dreblin-like protein; SH3 domain; cancer; diagnosis.  
XX  
OS Homo sapiens.  
XX  
PN JP2000197489-A.  
XX  
PD 18-JUL-2000.  
XX  
PF 07-JAN-1999; 99JP-00002254.  
XX  
PR 07-JAN-1999; 99JP-00002254.  
XX  
PA (KAGAKU) KAGAKU GIUTSU SHINKO JIGYODAN.  
XX  
DR WPI; 2000-545912/50.  
XX  
DR N-PSDB; AAA92163, AAA92164.  
XX  
PT New human protein having dreblin-like sequence and SH3 domain.  
XX  
PS Claim 1; Page 5-6; 12pp; Japanese.  
XX  
CC The present sequence represents a human protein having a dreblin-like  
CC sequence and SH3 domain. The protein and an antibody against the protein  
CC can be used in the diagnosis and the treatment of various diseases  
CC including cancers  
CC  
SQ Sequence 431 AA;

Query Match 20.0%; Score 796; DB 3; Length 431;  
Best Local Similarity 82.5%; Pred. No. 7.1e-73; Indels 0; Gaps 0;  
Matches 151; Conservative 2; Mismatches 30;

QY 359 RPERRPXRRPRTARSMGTWRTXSSISPPGKLRSPFLQXLTQTPETHFGREPAAXSR 418  
| | | | |  
DB 249 RNEQESAVHPREIFKQKERAMSTSSISPPGKLRSPFLQXLTQTPETHFGREPAAXSR 308  
| | | | |

QY 419 PRADLPAAEPAPSPPCLVQAEBAVYEEPEQSTTFEOPPLVQOQXGSEHIDHIIQOQ 478  
| | | | |  
DB 309 PRADLPAAEPAPSPPCLVQAEBAVYEEPEQSTTFEOPPLVQOQXGSEHIDHIIQOQ 368  
| | | | |

QY 479 GLSXQGLCARALYDQAAADXEISFDPENLITGIEVXXEGMWRGYPGDGHFGKXPANYVE 538  
| | | | |  
DB 369 GLSXQGLCARALYDQAAADXEISFDPENLITGIEVXXEGMWRGYPGDGHFGKXPANYVE 428  
| | | | |

QY 539 LIE 541  
| | | | |  
DB 429 LIE 431  
| | | | |

RESULT 10  
AAB85662  
ID AAB85662 standard; protein; 431 AA.  
XX  
AC AAB85662;  
XX  
DT 12-FEB-2001 (first entry)  
XX  
DE Human tyrosine kinase substrate cks118/Dresh protein sequence.  
XX  
KW Tyrosine kinase substrate; TKs; TKs 107; GRUB; TKs 113; TKs 118; Dresh;  
KW TKs 202; Src; rheumatoid arthritis; atherosclerosis; stroke; cancer;  
KW autoimmune disorder; organ transplantation; myocardial infarction;  
KW cardiomyopathy; renal failure; neurodegenerative disorder;  
KW cardiovascular disorder; inflammatory bowel disease; multiple sclerosis;  
KW asthma; osteoarthritis; psoriasis; rhinitis.  
XX  
OS Homo sapiens.  
XX  
PN WO200061750-A2.  
XX  
PD 19-OCT-2000.  
XX  
PF 06-APR-2000; 2000WO-US009277.  
XX  
PR 09-APR-1999; 99US-0128492P.  
XX  
PA (SUGEN-) SUGEN INC.  
XX  
PI Phan H, Courtnielde SA;  
XX  
DR WPI; 2000-679486/66.  
DR N-PSDB; AAC66078.  
XX  
PT Novel tyrosine kinase substrate polypeptides and polynucleotides, used to  
PT diagnose and treat diseases such as rheumatoid arthritis,  
PT atherosclerosis and cancer.  
XX  
PS Claim 2; Fig 7; 120pp; English.  
XX  
CC This invention relates to isolated, enriched or purified nucleic acid  
CC molecules represented by sequences AAC66076-66079 which encode tyrosine  
CC kinase substrate (TKs) TKs 107/GRUB, TKs 113, TKs 118/Dresh or TKs 202  
CC represented by sequences AAB85660-Y85663. The TKs proteins are substrates  
CC for cytoplasmic tyrosine kinase Src. The invention encompasses probe  
CC sequences, recombinant cells, antibodies specific for the TKs proteins,  
CC and hybridomas producing the antibodies. The TKs proteins exhibit  
CC antineumatic, antiarthritic, antiarteriosclerotic; immunosuppressive;  
CC antiinflammatory; osteopathic; antipsoriatic; antiallergic; and  
CC treat and diagnose diseases or disorders selected from rheumatoid  
CC arthritis, atherosclerosis, autoimmune disorders, organ transplantation,  
CC myocardial infarction, cardiomyopathy, stroke, renal failure, oxidative  
CC stress-related neurodegenerative disorders, cardiovascular disorders,  
CC cancer, and immune-related disorders selected from chronic inflammatory  
CC bowel disease, chronic inflammatory peptic disease, multiple sclerosis,  
CC asthma, osteoarthritis, psoriasis, and rhinitis. The proteins encoding the  
CC used to identify modulators of its activity. DNA sequences encoding the  
CC proteins may be used in gene therapy methods, and to produce transgenic  
CC animals  
CC  
SQ Sequence 431 AA;

Query Match 20.0%; Score 796; DB 3; Length 431;  
Best Local Similarity 82.5%; Pred. No. 7.1e-73; Indels 0; Gaps 0;  
Matches 151; Conservative 2; Mismatches 30;

QY RPERPPXXRPSRTASWGTWXTXSISSPOGKLRSPFLQXLOLPETHFGREPAAXSR 418  
 249 RNEQESAVHPREIRIFQKRAMSTSISSPOGKLRSPFLQXLOLPETHFGREPAALSR 308  
 QY 419 PRADLPAPBPAPSPPCLVQAEAEAYVEEPEQETFFYEQPPLVQOQXGSEHIDHHIOGQ 478  
 309 PRADLPAPBPAPSPPCLVQAEAEAYVEEPEQETFFYEQPPLVQOQXGSEHIDHHIOGQ 368  
 QY 479 GLSXQGLCARALYDQADDDXISFDPENLITGLEVXXEGMMRGYGPDPGKXPANYVE 538  
 369 GLSXQGLCARALYDQADDDXISFDPENLITGLEVXXEGMMRGYGPDPGKXPANYVE 428  
 QY 539 LIE 541  
 429 LIE 431

## RESULT 11

AAW79569 standard; protein; 458 AA.

AAW79569;

06-NOV-2001 (first entry)

Human protein SEQ ID NO 3215.

Human; cytokine; cell proliferation; cell differentiation; gene therapy; vaccine; peptide therapy; stem cell growth factor; haematopoiesis; tissue growth factor; immunomodulatory; cancer; leukaemia; nervous system disorder; arthritis; inflammation.

Homo sapiens.

MO200157190-A2.

09-AUG-2001.

05-FEB-2001; 2001WO-US004098.

03-FEB-2000; 2000US-00496914.  
 27-APR-2000; 2000US-00560875.  
 20-JUN-2000; 2000US-00598075.  
 19-JUL-2000; 2000US-00620325.  
 01-SEP-2000; 2000US-00654936.  
 15-SEP-2000; 2000US-00663561.  
 20-OCT-2000; 2000US-00693325.  
 30-NOV-2000; 2000US-00728422.

(HYSE-) HYSEQ INC.

Tang YT, Liu C, Dmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y, Zhao QA, Wang J, Zhang J, Ren F, Chen R, Wang ZM, Xue H, Yang Y, Wejman T, Goodrich R;

WPI; 2001-476283/51.

N-PSDB; AAK52702.

Nucleic acids encoding polypeptides with cytokine-like activities, useful in diagnosis and gene therapy.

Claim 20; Page 287; 6221pp: English.

The invention relates to polynucleotides (AAK51456-AAK53435) and the encoded polypeptides (AAW78323-AAW80302) that exhibit activity elating to cytokine, cell proliferation or cell differentiation or which may induce production of other cytokines in other cell populations. The polynucleotides and polypeptides are useful in gene therapy, vaccines or peptide therapy. The polypeptides have various cytokine-like activities, e.g. stem cell growth factor activity, haematopoiesis regulating activity, tissue growth factor activity, immunomodulatory activity and activin/inhibin activity and may be useful in the diagnosis and/or

CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
 CC inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111  
 CC (AAK52582) and 3666 (AAW80020) are omitted as the relevant pages from the  
 CC sequence listing were missing at the time of publication

XX Sequence 458 AA;

Query Match 20.0%; Score 796; DB 4; Length 458;

Best Local Similarity 82.5%; Pred. No. 7.7e-73; Matches 151; Conservative 2; Mismatches 30; Indels 0; Gaps 0;

QY 359 RPERPPXXRPSRTASWGTWXTXSISSPOGKLRSPFLQXLOLPETHFGREPAAXSR 418  
 276 RNEQESAVHPREIRIFQKRAMSTSISSPOGKLRSPFLQXLOLPETHFGREPAALSR 335  
 Db 419 PRADLPAPBPAPSPPCLVQAEAEAYVEEPEQETFFYEQPPLVQOQXGSEHIDHHIOGQ 478  
 336 PRADLPAPBPAPSPPCLVQAEAEAYVEEPEQETFFYEQPPLVQOQXGSEHIDHHIOGQ 395

QY 479 GLSXQGLCARALYDQADDDXISFDPENLITGLEVXXEGMMRGYGPDPGKXPANYVE 538  
 369 GLSXQGLCARALYDQADDDXISFDPENLITGLEVXXEGMMRGYGPDPGKXPANYVE 455  
 Db 539 LIE 541  
 456 LIE 458

QY 539 LIE 541  
 456 LIE 458

## RESULT 12

ABP98850 standard; protein; 377 AA.

ABP98850;

15-JUL-2003 (first entry)

Human structural and cytoskeletal associated protein #41.

Cytoskeletal; antiarteriosclerotic; anticonvulsant; nootropic; antiangiatic; neuroprotective; cerebroprotective; hypotensive; cardiant; osteopathic; antiinflammatory; antiarthritic; vincidine; gene therapy; human; structural and cytoskeleton-associated protein; SCAP; cancer; angina; atherosclerosis; epilepsy; Huntington's disease; hypertension; heart failure; osteoporosis; osteoarthritis.

Homo sapiens.

MO2003031940-A2.

17-APR-2003.

10-OCT-2002; 2002WO-US032851.

12-OCT-2001; 2001US-0328931P.  
 19-OCT-2001; 2001US-0360681P.  
 02-NOV-2001; 2001US-0343896P.  
 09-NOV-2001; 2001US-0346308P.  
 16-NOV-2001; 2001US-0332385P.  
 07-DEC-2001; 2001US-0340776P.  
 11-JAN-2002; 2002US-0347703P.

(INCY-) INCYTE GENOMICS INC.

Becha SD, Bhatia U, Blake JF, Borowsky ML, Burrill JD, Chang H, Chavala NK, Elliott VS, Emerling BM, Forsythe TJ, Gorrard AE, Lal PG, Griffin JA, Hafalia AJA, Ho A, Ison CH, Kable AE, Khare R, Lal PG, Lee S, Lee SA, Lee SY, Lehr-Watson PM, Li JX, Lindquist EA, Luo W, Marquis JP, Ramkumar J, Richardson TW, Sprague WW, Swarnakar A, Tang YT, Warren BA, Yang J, Yue H, Zabarjadian Y, Zheng W;

WPI; 2003-403125/38.  
 N-PSDB; ACC44337.



PT New human structural and cytoskeleton-associated proteins (SCAP) useful  
 for diagnosing, treating and preventing diseases or conditions associated  
 with the aberrant SCAP expression e.g. cancer, osteoporosis, or epilepsy.  
 PS Claim 1; Page 297; 361pp; English.

XX This sequence represents a novel isolated human structural and  
 CC cytoskeleton-associated protein (SCAP) polypeptide. The polypeptides and  
 CC polynucleotides encoding them are useful in diagnosing, treating and  
 CC preventing diseases or conditions associated with the decreased  
 CC expression or over expression of SCAP, such as cell proliferative (e.g.  
 CC cancer, atherosclerosis), neurological (e.g. epilepsy, Huntington's  
 CC disease, stroke), heart (e.g. hypertension, heart failure, angina) and  
 CC skeletal muscle disorders (e.g. osteoporosis, osteoarthritis) or viral  
 CC infections. These are also useful in assessing the effects of exogenous  
 CC compounds on the expression of nucleic acid and amino acid sequences of  
 CC SCAP. The SCAP or its fragments are useful in screening compounds for  
 CC effectiveness as agonist or antagonist of the polypeptides, or in  
 CC altering the expression of the target polynucleotide and compounds that  
 CC specifically bind to or modulate the activity of the polypeptide. The  
 CC microarray is useful in monitoring or measuring protein-protein  
 CC interactions, drug-target interactions, and gene expression profiles  
 CC  
 XX Sequence 377 AA;

## Query Match

Best Local Similarity 91.9%; Score 794; DB 6; Length 377;  
 Matches 148; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 381 TTSISSPQPKLRSPFLQKQLTQPEHFGREPAAXSRPRADLPAREPASPXPCLVQAE 440  
 DB 217 TTSISSPQPKLRSPFLQKQLTQPEHFGREPAAXSRPRADLPAREPASPXPCLVQAE 276  
 QY 441 BEAVYEEPPQETFEPPPLVQOQXGSEHIDHIOGQSLXQGLCARLYDYQADDXE 500  
 DB 277 BEAVYEEPPQETFEPPPLVQOQXGSEHIDHIOGQSLXQGLCARLYDYQADDXE 336  
 QY 501 ISFDPENLITGIEVXXEGMWRGYPDGHFGKXPANYVELIE 541  
 DB 337 ISFDPENLITGIEVIDEGMWRGYPDGHFGKXPANYVELIE 377

## RESULT 13

AB93895 standard; protein; 439 AA.

XX AAB93895;

DT 26-JUN-2001 (first entry)

XX Human protein sequence SEQ ID NO:13840.

DE Human; primer; detection; diagnosis; antisense therapy; gene therapy.

XX Homo sapiens.

PN EPI074617-A2.

PD 07-FEB-2001.

PF 28-JUL-2000; 2000EP-00116126.

PR 29-JUL-1999; 99JP-00249036.

PR 27-AUG-1999; 99JP-00300253.

PR 11-JAN-2000; 2000JP-00118776.

PR 02-MAY-2000; 2000JP-00183767.

PR 09-JUN-2000; 2000JP-00241899.

XX (HELI-) HELIX RES INST.

XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J,

PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

XX

DR WPI; 2001-318749/34.

XX Primer sets for synthesizing polynucleotides, particularly the 5602 full-  
 PT length cDNAs defined in the specification, and for the detection and/or  
 PT diagnosis of the abnormality of the proteins encoded by the full-length  
 PT cDNAs.

PS Claim 8; SEQ ID NO 13840; 2537bp + Sequence listing; English.

XX The present invention describes primer sets for synthesizing 5602 full-  
 CC length cDNAs defined in the specification. Where a primer set comprises:  
 CC (a) an oligo-dT primer and an oligonucleotide complementary to the  
 CC complementary strand of a polynucleotide which comprises one of the 5602  
 CC nucleotide sequences defined in the specification; where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence; where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in the  
 CC specification. The primer sets can be used in antisense therapy and in  
 CC gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13133 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893  
 CC represent human amino acid sequences; and AAH13629 to AAH13632 represent  
 CC oligonucleotides, all of which are used in the exemplification of the  
 CC present invention  
 CC  
 XX Sequence 439 AA;

## Query Match

Best Local Similarity 91.9%; Score 794; DB 4; Length 439;  
 Matches 148; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 381 TTSISSPQPKLRSPFLQKQLTQPEHFGREPAAXSRPRADLPAREPASPXPCLVQAE 440  
 DB 279 TTSISSPQPKLRSPFLQKQLTQPEHFGREPAAXSRPRADLPAREPASPXPCLVQAE 338  
 QY 441 BEAVYEEPPQETFEPPPLVQOQXGSEHIDHIOGQSLXQGLCARLYDYQADDXE 500  
 DB 339 BEAVYEEPPQETFEPPPLVQOQXGSEHIDHIOGQSLXQGLCARLYDYQADDXE 398  
 QY 501 ISFDPENLITGIEVXXEGMWRGYPDGHFGKXPANYVELIE 541  
 DB 399 ISFDPENLITGIEVIDEGMWRGYPDGHFGKXPANYVELIE 439

## RESULT 14

ABUS2367 standard; protein; 525 AA.

XX ABUS2367;

DT 03-MAR-2003 (first entry)

XX Human GPCR related protein NOV22a.

DE Human; NOVX; G-protein coupled receptor; GPCR; cancer; cytostatic.

XX Homo sapiens.

PN WO200279398-A2.

PD 10-OCT-2002.

PF 08-MAR-2002; 2002WO-US007355.

PR 08-MAR-2001; 2001US-0274194P.

PR 08-MAR-2001; 2001US-0274281P.

PR 08-MAR-2001; 2001US-0274322P.  
 PR 09-MAR-2001; 2001US-0274849P.  
 PR 13-MAR-2001; 2001US-0275578P.  
 PR 13-MAR-2001; 2001US-0275579P.  
 PR 13-MAR-2001; 2001US-0275601P.  
 PR 14-MAR-2001; 2001US-0276000P.  
 PR 16-MAR-2001; 2001US-0276776P.  
 PR 19-MAR-2001; 2001US-0276994P.  
 PR 20-MAR-2001; 2001US-0277239P.  
 PR 20-MAR-2001; 2001US-0277332P.  
 PR 21-MAR-2001; 2001US-0277791P.  
 PR 22-MAR-2001; 2001US-0278333P.  
 PR 23-MAR-2001; 2001US-0278152P.  
 PR 26-MAR-2001; 2001US-0278894P.  
 PR 27-MAR-2001; 2001US-0278999P.  
 PR 27-MAR-2001; 2001US-0279036P.  
 PR 30-MAR-2001; 2001US-0280233P.  
 PR 02-APR-2001; 2001US-0280802P.  
 PR 02-MAY-2001; 2001US-0280529P.  
 PR 02-MAY-2001; 2001US-0280666P.  
 PR 02-MAY-2001; 2001US-0288228P.  
 PR 17-MAY-2001; 2001US-0291766P.  
 PR 07-JUN-2001; 2001US-0296693P.  
 PR 08-JUN-2001; 2001US-0296856P.  
 PR 08-JUL-2001; 2001US-0303230P.  
 PR 05-JUL-2001; 2001US-0303237P.  
 PR 08-AUG-2001; 2001US-0310913P.  
 PR 13-AUG-2001; 2001US-0311978P.  
 PR 14-AUG-2001; 2001US-0312191P.  
 PR 16-AUG-2001; 2001US-0312916P.  
 PR 17-AUG-2001; 2001US-0313182P.  
 PR 20-AUG-2001; 2001US-0313626P.  
 PR 21-AUG-2001; 2001US-0314018P.  
 PR 27-AUG-2001; 2001US-0315227P.  
 PR 10-SEP-2001; 2001US-0318403P.  
 PR 10-SEP-2001; 2001US-0318510P.  
 PR 14-SEP-2001; 2001US-0322969P.  
 PR 14-SEP-2001; 2001US-0322360P.  
 PR 27-SEP-2001; 2001US-0325378P.  
 PR 09-NOV-2001; 2001US-0332486P.  
 PR 09-NOV-2001; 2001US-0345359P.  
 PR 07-MAR-2002; 2002US-00094886.

(CURA-) CURAGEN CORP.

PA Kekuda R, Tchernov VT, Liu X, Spytek KA, Patumraj M, Boldog FL,  
 PI Burgess CE, Vernet CAM, Li L, Gorman L, Malyankar UM, Miller CE, Casman SJ,  
 PI Guo X, Shenoy SG, Padigaru M, Taupier RJ, Smithson G, Zerhusen BD, Gerlach V,  
 PI Pena CE, Gangoli EA, Gusev V, Smithson G, Zerhusen BD, Gerlach V,  
 PI Pochart PF, Fernandes ER, Shinkets RA, Raselli L, Spaderna SK,  
 PI Larchelle WJ, Zhong M, Khramtsov NV, Voss EZ, Herrmann UJ,  
 XX WPI; 2003-058423/05.  
 DR N-PSDB; ABX70458.

PT NOVA polypeptides and polynucleotides, useful for treating a syndrome  
 related to a human disease associated with the NOVA polypeptide e.g.,  
 cancer.

PS Claim 1; Page 156; 413pp; English.

XX The present invention relates to the isolation of novel human  
 CC polypeptides referred to as NOVA (NOV1-NOV44), variants of these  
 CC proteins, and the polynucleotide sequences encoding them. The NOVA  
 CC proteins of the invention are G-protein coupled receptor (GPCR) related  
 CC proteins. The sequences of the invention are useful in the manufacture of  
 CC a medicament for treating a syndrome related to a human disease  
 CC associated with the polypeptides e.g. cancer. ABUS2311-ABUS2408 represent  
 CC the human NOVA proteins of the invention

XX Sequence 525 AA;

Query Match 18.7%; Score 746.5; DB 6; Length 525;  
 Best Local Similarity 67.3%; Pred. No. 1.2e-67;  
 Matches 148; Conservative 9; Mismatches 34; Indels 29; Gaps 3;

QY 381 TSSISPPQKURSPFLQXLTPTPTGRRPAXXSRRPDLPAEPPAPSPCLVQAE 440  
 DB 304 TTSISPPQKURSPFLQXLTPTPTGRRPAXXSRRPDLPAEPPAPSPCLVQAE 363  
 QY 441 EEAAYVEEPEQETFEQPPVLVQOQXGSEHIDHIIQGGSLSGQGLCARLYDYQADDE 500  
 DB 364 EEAAYVEEPEQETFEQPPVLVQOQXGSEHIDHIIQGGSLSGQGLCARLYDYQADDE 423  
 QY 501 ISFDPENLITGIEVXXEGMWRGYPDGHGKX-PANYVELIEX----- 542  
 DB 424 ISFDPENLITGIEVIDEGWRGYPDGHGKX-PANYVELIEX----- 542  
 QY 543 --GXGHILKFPQ-----TWLPYCKRRPXXHNAI 571  
 DB 484 RGGLGVDIQHSRRRTPSEDEASGLPAMQTPVTPNAAI 523

#### RESULT 15

AA73342  
 ID AAY73942 standard; protein; 162 AA.

AC AAY73942;

DT 14-MAR-2000 (first entry)

XX Human prostate tumor EST fragment derived protein #129.

DE Pancreas; tumor; EST; expressed sequence tag; human; cytostatic;

KW treatment.

OS Homo sapiens.

PN DE19820190-A1.

PD 04-NOV-1999.

XX 28-APR-1998; 98DE-01020190.

XX 28-APR-1998; 98DE-01020190.

XX (META-) METAGEN GES GENOMFORSCHUNG MBH.

PA Rosenthal A, Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E,  
 PI WPI; 1999-621386/54.  
 DR N-PSDB; AA252900.

PT New human nucleic acid sequences from pancreatic tumors, and related  
 proteins.

PS Claim 23; Page 363; 502pp; German.

XX This invention describes novel polypeptides and their encoding nucleic  
 CC acids derived from human pancreatic tumor tissue which have cytostatic  
 CC activity. The sequences are also useful in producing pharmaceutical  
 CC compositions for treatment of pancreatic tumors. AAY73814-Y74252  
 CC represent protein fragments encoded by the human pancreatic tumor cDNA  
 CC library derived expressed sequence tag (EST) sequences represented in  
 CC AA252858-253014

XX Sequence 162 AA;

Query Match 16.2%; Score 647; DB 2; Length 162;  
 Best Local Similarity 92.2%; Pred. No. 4.5e-58;  
 Matches 119; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 650 GASGRUWLPSPAFICLXFSLSASKGWPPLPFXKIGNSRRRLFLAEFTYXRVVDHGLIAA 709  
 DB 2 GASGRUWLPSPAFICLXFSLSASKGWPPLPFXKIGNSRRRLFLAEFTYXRVVDHGLIAA 61

Qy	710	GNLSCXLLCAPHSISLSICLGXKMGCRWPSSHPGSKXADTTGSTRRLTRCLQXVCA	769
Db	62	GNLSCWSLLCAPHSISLSICLGXKMGCRWPSSHPGSKXADTTGSTRRLTRCLQAPVCA	121
Qy	770	STDSDPRKS	778
Db	122	STDSDPRKS	130

Search completed: April 27, 2004, 10:52:56  
Job time : 55.8053 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2004, 10:51:05 ; Search time 20.1141 seconds

(without alignments)  
2035.354 Million cell updates/sec

Title: US-10-028-952A-9

Perfect score: 3989

Sequence: 1 HEIPVPTVYPAKQXXERA.....DFRKRKKKKLEKMTSSX 793

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A COMB pep.\*  
2: /cgn2\_6/ptodata/2/1aa/5B COMB pep.\*  
3: /cgn2\_6/ptodata/2/1aa/6A COMB pep.\*  
4: /cgn2\_6/ptodata/2/1aa/6B COMB pep.\*  
5: /cgn2\_6/ptodata/2/1aa/PCUS COMB pep.\*  
6: /cgn2\_6/ptodata/2/1aa/backfile1 pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	800.5	20.1	355	4	US-08-630-915A-192
2	628.5	15.8	433	4	US-08-630-915A-18
3	409	10.3	538	4	US-08-994-076-2
4	409	10.3	538	4	US-09-643-476-2
5	300	7.5	58	4	US-08-630-915A-118
6	300	7.5	58	4	US-08-630-915A-214
7	239.5	6.0	486	2	US-08-942-423-3
8	228.5	5.7	486	2	US-08-942-423-2
9	228.5	5.7	486	2	US-08-630-915A-26
10	208	5.2	546	2	US-08-942-423-4
11	208	5.2	546	2	US-08-630-915A-14
12	185.5	4.7	57	4	US-08-630-915A-116
13	185.5	4.5	53	2	US-08-942-423-37
14	178.5	4.5	57	4	US-08-630-915A-122
15	176.5	4.4	54	2	US-08-942-423-36
16	176.5	4.4	54	2	US-08-942-423-38
17	171.5	4.3	53	2	US-08-942-423-39
18	171.5	4.3	53	2	US-08-942-423-30
19	171.5	4.3	53	2	US-08-942-423-35
20	170.5	4.3	53	2	US-08-942-423-33
21	170.5	4.3	53	2	US-08-942-423-34
22	169.5	4.2	48	3	US-08-938-830-7
23	169.5	4.2	48	3	US-08-020-222-7
24	166.5	4.2	50	3	US-08-938-830-6
25	166.5	4.2	50	3	US-09-020-222-6
26	165.5	4.1	53	2	US-08-942-423-31
27	165.5	4.1	53	2	US-08-942-423-32

28	156.5	3.9	53	2	US-08-942-423-39	Sequence 39, Appl
29	148	3.7	205	4	US-08-630-915A-16	Sequence 16, Appl
30	148	3.7	401	2	US-08-549-004A-5	Sequence 5, Appl
31	148	3.7	401	3	US-09-051-982A-5	Sequence 5, Appl
32	141	3.5	261	2	US-08-691-814B-4	Sequence 4, Appl
33	138.5	3.5	456	4	US-09-252-991A-17335	Sequence 17335, A
34	130.5	3.3	497	4	US-09-252-991A-23620	Sequence 23620, A
35	127	3.2	696	3	US-08-906-865-4	Sequence 4, Appl
36	127	3.2	696	4	US-09-129-668-4	Sequence 4, Appl
37	125.5	3.1	1706	4	US-09-252-991A-31760	Sequence 31760, A
38	123	3.1	368	4	US-08-630-915A-20	Sequence 20, Appl
39	122	3.1	123	4	US-09-107-532A-5583	Sequence 5583, Ap
40	122	3.1	123	4	US-09-107-532A-5584	Sequence 5584, Ap
41	122	3.1	123	4	US-09-134-000C-6555	Sequence 6555, Ap
42	121	3.0	181	4	US-09-252-991A-26910	Sequence 26910, A
43	120.5	3.0	59	4	US-08-630-915A-117	Sequence 117, App
44	120.5	3.0	78	4	US-09-621-976-7727	Sequence 7727, Ap
45	119.5	3.0	802	4	US-09-252-991A-25050	Sequence 25050, A

## ALIGNMENTS

RESULT 1  
US-08-630-915A-192  
Sequence 192, Application US/08630915A  
Patent No. 6309820  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: HOFFMAN, No. 6309820h  
APPLICANT: KAY, Brian K.  
APPLICANT: FOWLES, Dana M.  
APPLICANT: MCCONNELL, Stephen J.  
TITLES OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
NUMBER OF SEQUENCES: 227  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/630,915A  
FILING DATE: 03-APR-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-174  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 192:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 355 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLTYPE: peptide  
US-08-630-915A-192  
Query Match 20.1%; Score 800.5; DB 4; Length 355;  
Best Local Similarity 64.9%; Pred. No. 5.3e-75;  
Matches 157; Conservative 9; Mismatches 47; Indels 29; Gaps 1;

QY 359 RPERPKXRPRRTASMGWTWXTXSSISPOGKLRSPHLOXOLTOPETHFEREPAXXSR 418  
 Db 112 RNEQSAVHREIFRQKRAMSTJSSISPOGKLRSPHLOXOLTOPETHFEREPAXXSR 171  
 QY 419 PRADPAEPAESPXPCUVOAEBAVYBEPXOEFTYQPPVQOQXGSEHIDHIQOQ 478  
 Db 172 PRADPAEPAESPXPCUVOAEBAVYBEPXOEFTYQPPVQOQXGSEHIDHIQOQ 231  
 QY 479 GLSXGLCARALYDQADDXEISFDPENLITGIEVXXGCMRGYGPDPGEXGKXPANYE 538  
 Db 232 GLSXGLCARALYDQADDXEISFDPENLITGIEVXXGCMRGYGPDPGEXGKXPANYE 291  
 QY 539 LIEGXGHLXFP-----SOTWLPYCWKRRPXXXXS 569  
 Db 292 LIDEAEGTSCPSPLAHGFLIAGRGGLVDIOHSSNRTPSDEASGLPFAWOTQPTVNA 351  
 QY 570 AL 571  
 Db 352 AM 353

## RESULT 2

US-08-630-915A-18  
 / Sequence 18, Application US/08630915A  
 / Patent No. 6309820  
 / GENERAL INFORMATION:  
 / APPLICANT: SPARKS, Andrew B.  
 / APPLICANT: HOFEMAN, No. 6309820h  
 / APPLICANT: KAY, Brian K.  
 / APPLICANT: FOMIKES, Dana M.  
 / APPLICANT: MCCONNELL, Stephen J.  
 / TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
 / TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
 / TITLE OF INVENTION: USING SAME  
 / NUMBER OF SEQUENCES: 227  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: Pennie & Edmonds LLP  
 / STREET: 1155 Avenue of the Americas  
 / CITY: New York  
 / STATE: New York  
 / COUNTRY: USA  
 / ZIP: 10036-2711  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: Floppy disk  
 / COMPUTER: IBM PC compatible  
 / OPERATING SYSTEM: PC-DOS/MS-DOS  
 / SOFTWARE: Patent in Release #1.0, Version #1.30  
 / CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US/08/630,915A  
 / FILING DATE: 03-Apr-1996  
 / CLASSIFICATION: 536  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: Mirock, S. Leslie  
 / REGISTRATION NUMBER: 18, 872  
 / REFERENCE/DOCKET NUMBER: 1101-174  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: (212) 790-9090  
 / TELEFAX: (212) 869-8864/9741  
 / TELEX: 66141 PENNIE  
 / INFORMATION FOR SEQ ID NO: 18:  
 / SEQUENCE CHARACTERISTICS:  
 / LENGTH: 433 amino acids  
 / TYPE: amino acid  
 / STRANDEDNESS:  
 / TOPOLOGY: unknown  
 / MOLECULE TYPE: peptide  
 / US-08-630-915A-18

Query Match 15.8%; Score 628.5; DB 4; Length 433;  
 Best Local Similarity 58.7%; Pred. No. 5.8e-57;  
 Matches 131; Conservative 12; Mismatches 43; Indels 37; Gaps 5;

QY 336 ARAPRAGTGVACVPEVLGNGARPERPKXRPRRTASMGWT-----W 378  
 Db 231 AGAPSRITG-----EPEQEAIV--SRTREWESAGQAPRPREIFQKERA 272  
 QY 379 RYTXSSISPOGKLRSPHLOXOLTOPETHFEREPAXXSRPRADLPAEPAESPXPCUVO 438  
 Db 273 MGTSTVTSOFGKLRSPHLOXOLTOPETHSYGRPTAPVSRPAAGV-CDEPASTLSS-AQ 330  
 QY 439 ABEBAVYBEPXOEFTYQPPVQOQXGSEHIDHIQOQGLSKQGLCARALYDQADD 498  
 Db 331 TEEEPYEVPEPDQTLTEEPPLVQOQAGSEHIDHYWQSGFGGGLCARALYDQADD 390  
 QY 499 XEISFDPENLITGIEVXXGCMRGYGPDPGEXGKXPANYELE 541  
 Db 391 TEISFDPENLITGIEVXXGCMRGYGPDPGEXGKXPANYELE 433

## RESULT 3

US-08-994-076-2  
 / Sequence 2, Application US/08994076  
 / Patent No. 6500937  
 / GENERAL INFORMATION:  
 / APPLICANT: Eryin, Jr., Paul R.  
 / TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCE  
 / TITLE OF INVENTION: OF NANASTATIN AND METHODS OF USE  
 / NUMBER OF SEQUENCES: 8  
 / CORRESPONDENCE ADDRESS:  
 / ADDRESSEE: Merchant, Gould, Smith, Edell, Welter & Schmidt  
 / STREET: 3100 No. 6500937west Center, 90 South 7th Street  
 / CITY: Minneapolis  
 / STATE: MN  
 / COUNTRY: USA  
 / ZIP: 55402  
 / COMPUTER READABLE FORM:  
 / MEDIUM TYPE: Diskette  
 / COMPUTER: IBM compatible  
 / OPERATING SYSTEM: DOS  
 / SOFTWARE: FastSeq for Windows Version 2.0  
 / CURRENT APPLICATION DATA:  
 / APPLICATION NUMBER: US/08/994,076  
 / FILING DATE: 19-DEC-1997  
 / CLASSIFICATION:  
 / PRIOR APPLICATION DATA:  
 / APPLICATION NUMBER: 60/027,315  
 / FILING DATE: 03-OCT-1996  
 / APPLICATION NUMBER: PCT/US97/18026  
 / FILING DATE: 03-OCT-1997  
 / APPLICATION NUMBER: 08/943,828  
 / FILING DATE: 03-OCT-1997  
 / ATTORNEY/AGENT INFORMATION:  
 / NAME: Kettelberger, Denise M  
 / REGISTRATION NUMBER: 33,924  
 / REFERENCE/DOCKET NUMBER: 4273.1US11  
 / TELECOMMUNICATION INFORMATION:  
 / TELEPHONE: 612-332-5300  
 / TELEFAX: 612-332-9081  
 / TELEX:  
 / INFORMATION FOR SEQ ID NO: 2:  
 / SEQUENCE CHARACTERISTICS:  
 / LENGTH: 538 amino acids  
 / TYPE: amino acid  
 / STRANDEDNESS: single  
 / TOPOLOGY: linear  
 / MOLECULE TYPE: protein  
 / FRAGMENT TYPE: internal  
 / US-08-994-076-2

Query Match 10.3%; Score 409; DB 4; Length 538;  
 Best Local Similarity 41.0%; Pred. No. 5.9e-34;  
 Matches 109; Conservative 11; Mismatches 62; Indels 84; Gaps 8;  
 QY 155 LILFSTNITDXXSASASRFXFGFXAGVKKVTITGILGWRPSVXSVAVFXSFVGSSTH 214  
 :|||||||:|||||

Db 1 MILFSTNTDRESGAS-----DLTFW-----VLSRRCQKSYH 33  
QY 215 XXAEFTKRWIV-----HPLIG---NXSWDXTVRVQVS-- 243  
Db 34 -----RDNWLVAAKRSKRRRFLILRCRLFLSLGSRIHQALDOSPNNREREGLDRRETL 88  
QY 244 FTLLMXCCCHGNPAQYERNRRXXHLYVYLGGANGAKLXSVGLXLNASKSESRRPGXTIRQ 303  
Db 89 FTLLMXCCCHGNPAQYERNRRFRHLVYVLGGANGAKLPSVGLLNASKSESRRPGXTIRQ 148  
QY 304 RRGASVGLGXPPXXRLSPAPGRPLHAPRXGRARAPPAAGTGVRCXVPFVLGNGARPER 363  
Db 149 RRGASVGLGXMPVPRLSPPAGAP-----PPAPRA-----REGACPAAR 187  
QY 364 PXXRPSRTARSWGTRXTXSISSPQ 389  
Db 188 RDGPVRSALRPGKRGAGKAAPSP 213

## RESULT 4

US-09-643-476-2  
Sequence 2, Application US/09643476  
Patent No. 6599495

## GENERAL INFORMATION:

APPLICANT: Evin, Jr., Paul R.  
TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCE  
OF MAMMASTATIN AND METHODS OF USE  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Merchant & Gould P.C.  
STREET: P.O. Box 2903  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55402-0903  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/643,476  
FILING DATE: 22-Aug-2000  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/027,315  
FILING DATE: 03-OCT-1996  
APPLICATION NUMBER: PCT/US97/18026  
FILING DATE: 03-OCT-1997  
APPLICATION NUMBER: 08/943,828  
FILING DATE: 03-OCT-1997  
APPLICATION NUMBER: 08/994,076  
FILING DATE: 19-DEC-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: Weaver, Karrie G.  
REGISTRATION NUMBER: 43,245  
REFERENCE/DOCKET NUMBER: 4213.1USC2  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-332-5300  
TELEFAX: 612-332-9081  
TELEX: <Unknown>  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 538 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
SEQUENCE DESCRIPTION: SEQ ID NO: 2:  
US-09-643-476-2

Query Match

10.3%; Score 409; DB 4; Length 538;

Best Local Similarity 41.0%; Pred. No. 5,9e-34;  
Matches 109; Conservative .11; Mismatches 62; Indels 84; Gaps 6;

QY 155 LILFSTNTDXXGASRGSPFPFGYAGVRYKTTGITGLMRSVXSDVAFXYDQSSYH 214  
Db 1 MILFSTNTDRESGAS-----DLTFW-----VLSRRCQKSYH 33  
QY 215 XXAEFTKRWIV-----HPLIG---NXSWDXTVRVQVS-- 243  
Db 34 -----RDNWLVAAKRSKRRRFLILRCRLFLSLGSRIHQALDOSPNNREREGLDRRETL 88  
QY 244 FTLLMXCCCHGNPAQYERNRRXXHLYVYLGGANGAKLXSVGLXLNASKSESRRPGXTIRQ 303  
Db 89 FTLLMXCCCHGNPAQYERNRRFRHLVYVLGGANGAKLPSVGLLNASKSESRRPGXTIRQ 148  
QY 304 RRGASVGLGXPPXXRLSPAPGRPLHAPRXGRARAPPAAGTGVRCXVPFVLGNGARPER 363  
Db 149 RRGASVGLGXMPVPRLSPPAGAP-----PPAPRA-----REGACPAAR 187  
QY 364 PXXRPSRTARSWGTRXTXSISSPQ 389  
Db 188 RDGPVRSALRPGKRGAGKAAPSP 213

## RESULT 5

US-08-630-915A-118  
Sequence 118, Application US/08630915A  
Patent No. 6309820

## GENERAL INFORMATION:

APPLICANT: HOFFMAN, Andrew B.  
APPLICANT: KAY, Brian K.  
APPLICANT: FOWLER, Dana M.  
APPLICANT: MCCONNELL, Stephen J.  
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
NUMBER OF SEQUENCES: 227  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/630,915A  
FILING DATE: 03-APR-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Mastrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-174  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/5741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 118:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 58 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-630-915A-118

Query Match

7.5%; Score 300; DB 4; Length 58;  
Best Local Similarity 93.1%; Pred. No. 6.9e-24;





```

CORRESPONDENCE ADDRESS:
ADDRESS: Syntex (U.S.A.) Inc.
STREET: 3401 Hillview Ave.
CITY: Palo Alto
STATE: California
COUNTRY: U.S.A.
ZIP: 94303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/942,423
FILING DATE: 01-OCT-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/362,715
FILING DATE: 23-DEC-1994
ATTORNEY/AGENT INFORMATION:
NAME: Perles, Rohan
REGISTRATION NUMBER: 35,752
REFERENCE/DOCKET NUMBER: 28260
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 852-1698
TELEFAX: (415) 496-3529
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: N-terminal
ORIGINAL SOURCE:
ORGANISM: HSI
US-08-942-423-2

Query Match      5.7%; Score 228.5; DB 2; Length 486;
Best Local Similarity 29.5%; Pred. No. 3.3e-15;
Matches 85; Conservative 24; Mismatches 122; Indels 57; Gaps 11;

QY 275 GANG-AKLXSVGLKLNASKESRPGXTIRQ--RGASVGLGXPPXRLSPAGRP--PLH 328
DB 234 GARGLKAFESLAERKREERKAQQVARKQOERKAVVMSREVQOPMPVEEPAAQAQ 293
QY 329 APRXRGARAPPRAGTGVACVPTVLGNGARPERRPXKPSRTARSMTGRTXTSISPO 388
DB 294 LPKTISSEVWP-----AESHLPPESQPVRSRREY-----PV 325
QY 389 PG-KLRSPFLQXOLQOPETHFERPAAXSR-PRADLPAEBPABSPCLVQAEBAAYE 446
DB 326 PSLPTQSPLOQNLHEDNE-----EPBALPRTPEGLQVVEEVEYAABEL-EPREPDYE 379
QY 447 EPXQETFEQPPVVOQXGSEHIDNH-----QGGLSXQGLCARALYD 492
DB 380 PEPEPTPEYEDVGEIDRQDEDAEGDYEDVLEPEDTPSLSYQAGPSAGAGASIALYD 439
QY 493 YQAADDEXISFDPENLITGIEVXXEGWWRGYGPDGHFGKXPANYVELI 540
DB 440 YQGGSDLSFDPDITIDIEVWDEGWRGQ-CRGHFGLFPANYVKLL 486

```

RESULT 9  
 US-08-630-915A-26  
 Sequence 26, Application US/08630915A  
 Patent No. 6309820  
 GENERAL INFORMATION:  
 APPLICANT: SPARKS, Andrew B.  
 APPLICANT: HOFFMAN, NO. 6309820H  
 APPLICANT: KAY, Brian K.

```

APPLICANT: FOWLES, Dana M.
APPLICANT: MCCONNELL, Stephen J.
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND
TITLE OF INVENTION: USING SAME
NUMBER OF SEQUENCES: 227
CORRESPONDENCE ADDRESS:
ADDRESS: Pennie & Edmonds LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,915A
FILING DATE: 03-APR-1996
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Mastrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 1101-174
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/5741
TEXT: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 26:
SEQUENCE CHARACTERISTICS:
LENGTH: 486 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: unknown
MOLECULE TYPE: peptide
US-08-630-915A-26

Query Match      5.7%; Score 228.5; DB 4; Length 486;
Best Local Similarity 29.5%; Pred. No. 3.3e-15;
Matches 85; Conservative 24; Mismatches 122; Indels 57; Gaps 11;

QY 275 GANG-AKLXSVGLKLNASKESRPGXTIRQ--RGASVGLGXPPXRLSPAGRP--PLH 328
DB 234 GARGLKAFESLAERKREERKAQQVARKQOERKAVVMSREVQOPMPVEEPAAQAQ 293
QY 329 APRXRGARAPPRAGTGVACVPTVLGNGARPERRPXKPSRTARSMTGRTXTSISPO 388
DB 294 LPKTISSEVWP-----AESHLPPESQPVRSRREY-----PV 325
QY 389 PG-KLRSPFLQXOLQOPETHFERPAAXSR-PRADLPAEBPABSPCLVQAEBAAYE 446
DB 326 PSLPTQSPLOQNLHEDNE-----EPBALPRTPEGLQVVEEVEYAABEL-EPREPDYE 379
QY 447 EPXQETFEQPPVVOQXGSEHIDNH-----QGGLSXQGLCARALYD 492
DB 380 PEPEPTPEYEDVGEIDRQDEDAEGDYEDVLEPEDTPSLSYQAGPSAGAGASIALYD 439
QY 493 YQAADDEXISFDPENLITGIEVXXEGWWRGYGPDGHFGKXPANYVELI 540
DB 440 YQGGSDLSFDPDITIDIEVWDEGWRGQ-CRGHFGLFPANYVKLL 486

```

RESULT 10  
 US-08-942-423-4  
 Sequence 4, Application US/08942423  
 Patent No. 5891673  
 GENERAL INFORMATION:  
 APPLICANT: Hashimoto, Yasuhiro  
 APPLICANT: Takemoto, Yoshihiro  
 TITLE OF INVENTION: Lock Binding Protein  
 NUMBER OF SEQUENCES: 68



SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/630,915A  
FILING DATE: 03-APR-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-174  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 116:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 57 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-630-915A-116

Query Match 4.7%; Score 185.5; DB 4; Length 57;  
Best Local Similarity 62.5%; Pred. No. 5.0e-12;  
Matches 35; Conservative 8; Mismatches 12; Indels 1; Gaps 1;

QY 484 GLCARALDYQAADDXISFPDENLITIGIEVXXEGMWRGYGPDGHFGKXPANYVEL 539  
DB 2 GISAIALVDYQAGSDLSFPDDITITIDIEVDEGMWRGQ-CRGHFGLFPANYVEL 56

RESULT 13  
US-08-942-423-37  
Sequence 37, Application US/08942423  
Patent No. 5891673  
GENERAL INFORMATION:  
APPLICANT: Hashimoto, Yasuhiro  
APPLICANT: Takemoto, Yoshihiro  
TITLE OF INVENTION: Lck Binding Protein  
NUMBER OF SEQUENCES: 68  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Syntex (U.S.A.) Inc.  
STREET: 3401 Hillyview Ave.  
CITY: Palo Alto  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 94303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/942,423  
FILING DATE: 01-OCT-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,715  
FILING DATE: 23-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Perles, Rohan  
REGISTRATION NUMBER: 35,752  
REFERENCE/DOCKET NUMBER: 28260  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 852-1698  
TELEFAX: (415) 496-3529  
INFORMATION FOR SEQ ID NO: 37:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 53 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide

US-08-942-423-37  
Query Match 4.5%; Score 180.5; DB 2; Length 53;  
Best Local Similarity 63.0%; Pred. No. 1.7e-11;  
Matches 34; Conservative 8; Mismatches 11; Indels 1; Gaps 1;

QY 487 AALVDYQAADDXISFPDENLITIGIEVXXEGMWRGYGPDGHFGKXPANYVELI 540  
DB 1 AALVDYQAAGSDLSFPDDITITIDIEVDEGMWRGQ-CRGHFGLFPANYVELL 53

RESULT 14  
US-08-630-915A-122  
Sequence 122, Application US/08630915A  
Patent No. 6309820  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
APPLICANT: HOFFMAN, No. 6309820h  
APPLICANT: KAY, Brian K.  
APPLICANT: FOLKES, Dana M.  
APPLICANT: MCCONNELL, Stephen J.  
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
TITLE OF INVENTION: DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
NUMBER OF SEQUENCES: 227  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/630,915A  
FILING DATE: 03-APR-1996  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistrock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-174  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 122:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 57 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-08-630-915A-122

Query Match 4.5%; Score 178.5; DB 4; Length 57;  
Best Local Similarity 57.9%; Pred. No. 3.1e-11;  
Matches 33; Conservative 9; Mismatches 14; Indels 1; Gaps 1;  
QY 484 GLCARALDYQAADDXISFPDENLITIGIEVXXEGMWRGYGPDGHFGKXPANYVELI 540  
DB 2 GISAIALVDYQAGSDLSFPDDITITIDIEVDEGMWRGQ-CRGHFGLFPANYVELL 57  
RESULT 15  
US-08-942-423-36  
Sequence 36, Application US/08942423  
Patent No. 5891673  
GENERAL INFORMATION:  
APPLICANT: Hashimoto, Yasuhiro

Wed Apr 28 09:57:51 2004

us-10-028-952a-9.rat

Page 8

APPLICANT: Takemoto, Yoshihiro  
TITLE OF INVENTION: Lock Binding Protein  
NUMBER OF SEQUENCES: 68  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Syntex (U.S.A.) Inc.  
STREET: 3401 Hillview Ave.  
CITY: Palo Alto  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 94303  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/942,423  
FILING DATE: 01-OCT-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/362,715  
FILING DATE: 23-DEC-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Perles, Rohan  
REGISTRATION NUMBER: 35,752  
REFERENCE/DOCKET NUMBER: 28260  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 852-1698  
TELEFAX: (415) 496-3529  
INFORMATION FOR SEQ ID NO: 36:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 54 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-942-423-36

Query Match 4.4%; Score 176.5; DB 2; Length 54;  
Best Local Similarity 60.0%; Pred. No. 4.7e-11;  
Matches 33; Conservative 8; Mismatches 13; Indels 1; Gaps 1;  
QY 487 ARALYDYOADADXEISFDPENLITGIEVXXEGWNRGSPDGHFGKXKPPATYVELIE 541  
DB 1 AIALYDYOQEGSDELSFDPDDITTDIEWVEGWRGQ-CRGHFGLFPATYKLE 54

Search completed: April 27, 2004, 10:56:53  
Job time: 21.1141 secs



```

1 OTHER INFORMATION: Xaa is any amino acid or may be absent
2 FEATURE:
3 NAME/KEY: UNSURE
4 LOCATION: (274)
5 OTHER INFORMATION: Xaa is any amino acid or may be absent
6 FEATURE:
7 NAME/KEY: UNSURE
8 LOCATION: (282)
9 OTHER INFORMATION: Xaa is any amino acid or may be absent
10 FEATURE:
11 NAME/KEY: UNSURE
12 LOCATION: (287)
13 OTHER INFORMATION: Xaa is any amino acid or may be absent
14 FEATURE:
15 NAME/KEY: UNSURE
16 LOCATION: (299)
17 OTHER INFORMATION: Xaa is any amino acid or may be absent
18 FEATURE:
19 NAME/KEY: UNSURE
20 LOCATION: (312)
21 OTHER INFORMATION: Xaa is any amino acid or may be absent
22 FEATURE:
23 NAME/KEY: UNSURE
24 LOCATION: (315) - (316)
25 OTHER INFORMATION: Xaa is any amino acid or may be absent
26 FEATURE:
27 NAME/KEY: UNSURE
28 LOCATION: (332)
29 OTHER INFORMATION: Xaa is any amino acid or may be absent
30 FEATURE:
31 NAME/KEY: UNSURE
32 LOCATION: (349)
33 OTHER INFORMATION: Xaa is any amino acid or may be absent
34 FEATURE:
35 NAME/KEY: UNSURE
36 LOCATION: (365) - (366)
37 OTHER INFORMATION: Xaa is any amino acid or may be absent
38 FEATURE:
39 NAME/KEY: UNSURE
40 LOCATION: (380)
41 OTHER INFORMATION: Xaa is any amino acid or may be absent
42 FEATURE:
43 NAME/KEY: UNSURE
44 LOCATION: (382)
45 OTHER INFORMATION: Xaa is any amino acid or may be absent
46 FEATURE:
47 NAME/KEY: UNSURE
48 LOCATION: (399)
49 OTHER INFORMATION: Xaa is any amino acid or may be absent
50 FEATURE:
51 NAME/KEY: UNSURE
52 LOCATION: (415) - (416)
53 OTHER INFORMATION: Xaa is any amino acid or may be absent
54 FEATURE:
55 NAME/KEY: UNSURE
56 LOCATION: (432)
57 OTHER INFORMATION: Xaa is any amino acid or may be absent
58 FEATURE:
59 NAME/KEY: UNSURE
60 LOCATION: (449)
61 OTHER INFORMATION: Xaa is any amino acid or may be absent
62 FEATURE:
63 NAME/KEY: UNSURE
64 LOCATION: (465) - (466)
65 OTHER INFORMATION: Xaa is any amino acid or may be absent
66 FEATURE:
67 NAME/KEY: UNSURE
68 LOCATION: (482)
69 OTHER INFORMATION: Xaa is any amino acid or may be absent
70 FEATURE:
71 NAME/KEY: UNSURE
72 LOCATION: (499)
73 OTHER INFORMATION: Xaa is any amino acid or may be absent

```

```

FEATURE:
NAME/KEY: UNSURE
LOCATION: (515)..(516)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (532)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (542)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (544)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE

```

Query Match 95.9%; Score 3824; DB 14; Length 793;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 792; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 HEIPVPTYYPAKPOXERAMENQKKTLLSTLVHGEETXEVXNMKAPGAPVSPRG 60
DB 1 HEIPVPTYYPAKPOXERAMENQKKTLLSTLVHGEETXEVXNMKAPGAPVSPRG 60
QY 61 ARGXXXRPGPPVVKTHYSRFTDPRRGGEPRGALAXARPARRGATRSCKXAMWG 120
DB 61 ARGXXXRPGPPVVKTHYSRFTDPRRGGEPRGALAXARPARRGATRSCKXAMWG 120
QY 121 VILGTYTCQTVQVEXGELREDRNLPMXRAKARILLFSTNTDXXSGASSFXPFGFXA 180
DB 121 VILGTYTCQTVQVEXGELREDRNLPMXRAKARILLFSTNTDXXSGASSFXPFGFXA 180
QY 181 GXVRKVTGTIGLWEPVSXSDVAFYSFDVGSYHXXAEFTKMTVHPLIGXSMXTVVR 240
DB 181 GXVRKVTGTIGLWEPVSXSDVAFYSFDVGSYHXXAEFTKMTVHPLIGXSMXTVVR 240
QY 241 QVSTFLMKCCCHGPAQYERXRRXXHLVYVVGXGANGAKLSVGLXNASKSESPGXT 300
DB 241 QVSTFLMKCCCHGPAQYERXRRXXHLVYVVGXGANGAKLSVGLXNASKSESPGXT 300
QY 301 IQRRGASVGLGXPXXRLSPAGRPPLHAPRYRGARAPPRAGTGVRXVPPVLGNGAR 360
DB 301 IQRRGASVGLGXPXXRLSPAGRPPLHAPRYRGARAPPRAGTGVRXVPPVLGNGAR 360
QY 361 ERPRXXRPSRTARSGTWRTXTSISPOGKLRSPFLQXOLQPTHRGRRPAXXSR 420
DB 361 ERPRXXRPSRTARSGTWRTXTSISPOGKLRSPFLQXOLQPTHRGRRPAXXSR 420
QY 421 ADLPAREPAPSPCIVQAEAEVYEBPXQETFEYQEPPLVQOQXGSEHDIHQOGL 480
DB 421 ADLPAREPAPSPCIVQAEAEVYEBPXQETFEYQEPPLVQOQXGSEHDIHQOGL 480
QY 481 SYQGLCARALYIOAADXELISFPDENLITGIEVXXEGEWMRGYGDGHFGXKPNYVLI 540
DB 481 SYQGLCARALYIOAADXELISFPDENLITGIEVXXEGEWMRGYGDGHFGXKPNYVLI 540
QY 541 EXGKGHILKFPSTQMLPYCMKRRPXXHSLFOEADPOXGKTLRAPSGLASACHPKCN 600
DB 541 EXGKGHILKFPSTQMLPYCMKRRPXXHSLFOEADPOXGKTLRAPSGLASACHPKCN 600
QY 601 GLVPTHPSCIPRPXXTAMLLPLTGYXAPCKWBPBPMPBSCGSGSXGASGRMLPSA 660
DB 601 GLVPTHPSCIPRPXXTAMLLPLTGYXAPCKWBPBPMPBSCGSGSXGASGRMLPSA 660
QY 661 FICLXXFSLASGWWPPLFRXKLGNSERRELFLAFVTVXVRDHDGLAAGNLSCKXLLCA 720
DB 661 FICLXXFSLASGWWPPLFRXKLGNSERRELFLAFVTVXVRDHDGLAAGNLSCKXLLCA 720
QY 721 PHSISLCLGXGKMGCRWPSHPGYSKXADTTCSTRLTRCLQXXVCASTDSPFRSKK 780
DB 721 PHSISLCLGXGKMGCRWPSHPGYSKXADTTCSTRLTRCLQXXVCASTDSPFRSKK 780

```

```

DB 721 PHSISLCLGXGKMGCRWPSHPGYSKXADTTCSTRLTRCLQXXVCASTDSPFRSKK 780
QY 781 KXKLEKLTSS 792
DB 781 KXKLEKLTSS 792

```

```

RESULT 2
US-10-028-952a-10
Sequence 10, Application US/10028952A
Publication No. US20030157576A1
GENERAL INFORMATION:
APPLICANT: Ervin, Jr., Paul R.
TITLE OF INVENTION: EPITHELIAL CELL GROWTH INHIBITORS
FILE REFERENCE: 4273.3USM1
CURRENT APPLICATION NUMBER: US/10/028, 952A
CURRENT FILING DATE: 2001-12-18
PRIOR APPLICATION NUMBER: PCT/US00/16900
PRIOR FILING DATE: 2000-06-19
PRIOR APPLICATION NUMBER: 60/139, 995
PRIOR FILING DATE: 1999-06-18
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 10
LENGTH: 784
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: UNSURE
LOCATION: (16)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (13)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (16)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (33)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (49)..(50)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (66)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (83)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (90)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (99)..(100)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (104)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (116)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE

```



LOCATION: (133)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (142)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (147)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (149) .. (150)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (166)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (173)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (183)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (199) .. (200)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (204)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (216)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (233)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (242)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (249) .. (250)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (255)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (266)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (281)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (283)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (299) .. (300)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (316)

OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (333)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (349) .. (350)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (366) .. (368)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (378)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (382) .. (383)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (386)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (398) .. (401)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (416)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (427)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (433)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (449) .. (450)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (456)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (466)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (483)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (499) .. (500)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE

Query Match 32.3%; Score 1288; DB 14; Length 784;  
Best Local Similarity 80.3%; Pred. No. 2.7e-104; Indels 4; Gaps 2;  
Matches 282; Conservative 7; Mismatches 58;

QY 37 HGEETXEYXNKMXPAPVSPRGARGXKXPCGPVYKTHYSDFXTDVPRGGEPRGAL 96  
DB 5 HGEETXEYXNKMXPAPVSPRGARGXKXPCGPVYKTHYSDFXTDVPRGGEPRGAL 64  
QY 97 ASXAKRPAAKPGATRGXKXKMGVILGRYTCQTVXQVSGXGELDENRILPWSKXAKRLI 156

Db 65 AKGAKRPAARPPGATRGSGXSAARMGVXLGRTYCCQTXOVASGEXLJEDNLFWMXRRAKALFI 124  
Qy 157 LIFSTNDXXSGARSFPFGFXGXYRKVTGTGLMRPVSXDVAFXSFVGSYHXX 216  
Db 125 LIFSTNDXSGARSFPFGFXGXYRKVTGTGLMRPVSXDVAFXSFVGSYHXX 184  
Qy 217 AEFTRMIVHPLIGXSWXVTYVQVSEFTLLMCCCHGNPAQYERNRXXHLVYLXGA 276  
Db 185 AEFTRMIVHPLIGXSWXVTYVQVSEFTLLMCCCHGNPAQYERNRXXHLVYLXGA 244  
Qy 277 NGAKLXSVGLXNLNKSSESXPRXGTRORGNASVGLGXPKYXRLSPAPRPLHAPRXGSA 336  
Db 245 NGAKKXSVGLXNLNKSSESXPRXGTRORGNASVGLGXPKYXRLSPAPRPPSTXXRAGG 304  
Qy 337 RAPPRA--GTGVRXCVFVLNGARPPRRPXXRPSRTASWGTWXTXSI 384  
Db 305 RVPRRAPPGSXACPSMSWETGRG-RKCGXPLAHAPHVBARAPFXXSSTI 354

## RESULT 3

US-09-879-957-192  
Sequence 192, Application US/09879957  
Patent No. US20020034755A1  
GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.  
HOFMAN, No. US20020034755A1h

KAY, Brian K.

FOMLES, Dana M.

MCCONNELL, Stephen J.

TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
USING SAME

NUMBER OF SEQUENCES: 227

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/879,957

FILING DATE: 13-Jun-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/630,915

FILING DATE: 03-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: Mistrock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 1101-174

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864/9741

TELEX: 66141 PENNMB

INFORMATION FOR SEQ ID NO: 192:

SEQUENCE CHARACTERISTICS:

LENGTH: 355 amino acids

TYPE: amino acid

STRANDEDNESS: <unknown>

TOPOLOGY: unknown

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 192:

US-09-879-957-192

Matches 157; Conservative 9; Mismatches 47; Indels 29; Gaps 1;  
Qy 359 RPRRPXKRPSTASWGTWXTXSISSPQPKLSPLQXLTQPTHTFGREPAAXSR 418  
Db 112 RNEQSAVHREIFQKERAAMSTSISSPQPKLSPLQXLTQPTHTFGREPAAXSR 171  
Qy 419 PRADLPAPRPSXPPCLVQAEBAVYEPKQEFYFQPPVQOQXGSHIDHIQOQ 478  
Db 172 PRADLPAPRPSXPPCLVQAEBAVYEPKQEFYFQPPVQOQXGSHIDHIQOQ 231  
Qy 479 GLSXGCLCARALYDQADDXEISFPENLITGIEVXXEGWWRGYGPDGHPGXPANYE 538  
Db 232 GLSGGCLCARALYDQADDXEISFPENLITGIEVXXEGWWRGYGPDGHPGXPANYE 291  
Qy 539 LIEKXGHTLXFP-----SQTWLYCWRKRPXXHS 569  
Db 292 LIDEAVGTCSPPLRNGFLIAGRGVLDIQHSNRTPSEDEASGLPPMCTQPTVPA 351  
Qy 570 AL 571  
Db 352 AM 353

## RESULT 4

US-10-094-886-114  
Sequence 114, Application US/10094886  
Publication No. US20040002120A1  
GENERAL INFORMATION:

APPLICANT: Kekuda, Ramesh

APPLICANT: Tchernev, Velizar T.

APPLICANT: Liu, Xiaohong

APPLICANT: Spytek, Kimberly A.

APPLICANT: Patnursajan, Meera

APPLICANT: Burgess, Catherine

APPLICANT: Verne, Corine A.

APPLICANT: Li, Li

APPLICANT: Goeman, Linda

APPLICANT: Malyanekar, Uriel M.

APPLICANT: Boldog, Ferenc

APPLICANT: Guo, Xiaojia

APPLICANT: Shenoy, Suresh

APPLICANT: Padigaru, Muralidhara

APPLICANT: Taupier, Raymond J., Jr.

APPLICANT: Miller, Charles

APPLICANT: Casman, Stacie

APPLICANT: Pena, Carol

APPLICANT: Gangoli, Esha

APPLICANT: Gusev, Vladimir

APPLICANT: Smithson, Glenda

APPLICANT: Zernusen, Bryan

APPLICANT: Gerlach, Valerie

APPLICANT: Pochart, Pascal

APPLICANT: Fernandes, Elma

APPLICANT: Shimkets, Richard

APPLICANT: Rastelli, Luca

APPLICANT: Spaderina, Steven

APPLICANT: Larochele, William

APPLICANT: Zhong, Mei

TITLE OF INVENTION: THERAPEUTIC POLYPEPTIDES, NUCLEIC ACIDS ENCODING SAME, AND METHOD

FILE REFERENCE: 21402-290 B

CURRENT APPLICATION NUMBER: US/10/094,886

CURRENT FILING DATE: 2002-03-07

PRIOR APPLICATION NUMBER: 60/274,322

PRIOR FILING DATE: 2001-03-08

PRIOR APPLICATION NUMBER: 60/313,182

PRIOR FILING DATE: 2001-08-17

PRIOR APPLICATION NUMBER: 60/288,052

PRIOR FILING DATE: 2001-05-02

PRIOR APPLICATION NUMBER: 60/318,510

PRIOR FILING DATE: 2001-09-10

PRIOR APPLICATION NUMBER: 60/274,281

PRIOR FILING DATE: 2001-03-08

PRIOR APPLICATION NUMBER: 60/314,018

Query Match 20.1%; Score 800.5; DB 9; Length 355;  
Best Local Similarity 64.9%; Pred. No. 7.3e-62;

Wed Apr 28 09:57:52 2004

us-10-028-952a-9.rapb

Page 6

PRIOR FILING DATE: 2001-08-21  
PRIOR APPLICATION NUMBER: 60/274,194  
PRIOR FILING DATE: 2001-03-08  
PRIOR APPLICATION NUMBER: 60/274,849  
PRIOR FILING DATE: 2001-03-09  
PRIOR APPLICATION NUMBER: 60/296,693  
PRIOR FILING DATE: 2001-06-07  
PRIOR APPLICATION NUMBER: 60/313,626  
Remaining Prior Application data removed - See File Wrapper or PALM.  
NUMBER OF SEQ ID NOS: 298  
SOFTWARE: Patent 2.1  
SEQ ID NO: 114  
LENGTH: 525  
TYPE: PRF  
ORGANISM: Homo sapiens  
US-10-094-886-114

Query Match 18.7%; Score 746.5; DB 15; Length 525;  
Best Local Similarity 67.3%; Pred. No. 6.3e-57;  
Matches 148; Conservative 9; Mismatches 34; Indels 29; Gaps 3;

QY 361 TYSISSPQGLRSPFLQXLTQPEHFGREPAAXSRPADLPAEPAASXPCLVOAE 440  
DB 304 TTSISSPQGLRSPFLQXLTQPEHFGREPAAXSRPADLPAEPAASXPCLVOAE 363  
QY 441 EEAAYEEXPQETFEYEOPLVQOQXGSEHIDHIIQCGLSXQGLCARALYDQADDXE 500  
DB 364 EEAAYEEXPQETFEYEOPLVQOQXGSEHIDHIIQCGLSXQGLCARALYDQADDXE 423  
QY 501 ISFDPENITGIEVYXEGMWRGYPGDHFGMX-PANVELIEX----- 542  
DB 424 ISFDPENITGIEVYXEGMWRGYPGDHFGMXPTTWSSEAGSCSPRLRGTLLNG 483  
QY 543 --GXGHTLXFPSSQ-----TWLPYCKTRPPXXHXSAL 571  
DB 484 RGGIGVDIQHSSRNRTSEDEASGLPPAWOTQPTVPAAM 523

RESULT 5

US-09-879-957-18  
Sequence 18, Application US/09879957  
Patent No. US20020034755A1  
GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.  
HOFFMAN, No. US20020034755A1h  
KAY, Brian K.  
FOWLES, Dana M.  
MCCONNELL, Stephen J.

TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
USING SAME

NUMBER OF SEQUENCES: 227

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/879,957  
FILING DATE: 13-Jun-2001  
CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/630,915  
FILING DATE: 03-APR-1996  
ATTORNEY/AGENT INFORMATION:

NAME: Mirock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-174  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 18:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 433 amino acids  
TYPE: amino acid  
STRANDEDNESS: <unknown>  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 18:  
US-09-879-957-18

Query Match 15.8%; Score 628.5; DB 9; Length 433;  
Best Local Similarity 58.7%; Pred. No. 1.2e-46;  
Matches 131; Conservative 12; Mismatches 43; Indels 37; Gaps 5;

QY 336 ARAPPAGTGVRCVFFVLGNGARPERRPXXSRTRASMG-----W 378  
DB 231 AGAPSRGT-----EPQZAV-----SKTQWESAGQOAPRRIFFQXERA 272  
QY 379 RXTYSISSPQGLRSPFLQXLTQPEHFGREPAAXSRPADLPAEPAASXPCLVO 438  
DB 273 MSTTSVSSQGLRSPFLQXLTQPEHFGREPAAXSRPADLPAEPAASXPCLVO 330  
QY 439 AEEAYEEXPQETFEYEOPLVQOQXGSEHIDHIIQCGLSXQGLCARALYDQADDXE 498  
DB 331 TEEPTFEVPEPDITVEPPLVQOQXGSEHIDHIIQCGLSXQGLCARALYDQADDXE 390  
QY 499 XEISFDPENITGIEVYXEGMWRGYPGDHFGMX-PANVELIE 541  
DB 391 TEISFDPENITGIEVYXEGMWRGYPGDHFGMX-PANVELIE 433

RESULT 6

US-09-879-957-118  
Sequence 118, Application US/09879957  
Patent No. US20020034755A1  
GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.  
HOFFMAN, No. US20020034755A1h  
KAY, Brian K.  
FOWLES, Dana M.  
MCCONNELL, Stephen J.

TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
USING SAME

NUMBER OF SEQUENCES: 227

CORRESPONDENCE ADDRESS:

ADDRESSEE: Pennie & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/879,957  
FILING DATE: 13-Jun-2001  
CLASSIFICATION: <unknown>

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/630,915  
FILING DATE: 03-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Mirock, S. Leslie

REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-174  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 118:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 58 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
SEQUENCE DESCRIPTION: SEQ ID NO: 118:  
US-09-879-957-118

Query Match 7.5%; Score 300; DB 9; Length 58;  
Best Local Similarity 93.1%; Pred. No. 8,6e-19;  
Matches 54; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 483 QGLCARALDYQADADDEISFDPENLITGIEVXXEGMWRGYPDGHFGKXPANYVELI 540  
Db 1 QGLCARALDYQADADDEISFDPENLITGIEVIDEGMWRGYPDGHFGKXPANYVELI 58

## RESULT 7

US-09-879-957-214  
Sequence 214, Application US/09879957  
Patent No. US20020034755A1  
GENERAL INFORMATION:

APPLICANT: SPARKS, Andrew B.  
HOFMAN, No. US20020034755A1h

KAY, Brian K.  
FOWLES, Dana M.

McCONNELL, Stephen J.  
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL

DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
USING SAME

NUMBER OF SEQUENCES: 227

CORRESPONDENCE ADDRESS:

ADDRESSES: Pennie & Edmonds LLP

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: USA

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/879,957

FILING DATE: 13-Jun-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/630,915

FILING DATE: 03-APR-1996

ATTORNEY/AGENT INFORMATION:

NAME: Mierock, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 1101-174

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 790-9090

TELEFAX: (212) 869-8864/9741

INFORMATION FOR SEQ ID NO: 214:

SEQUENCE CHARACTERISTICS:

LENGTH: 58 amino acids

TYPE: amino acid

STRANDEDNESS: <Unknown>

TOPOLOGY: unknown

MOLECULE TYPE: peptide

SEQUENCE DESCRIPTION: SEQ ID NO: 214:  
US-09-879-957-214

Query Match 7.5%; Score 300; DB 9; Length 58;  
Best Local Similarity 93.1%; Pred. No. 8,6e-19;  
Matches 54; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 483 QGLCARALDYQADADDEISFDPENLITGIEVXXEGMWRGYPDGHFGKXPANYVELI 540  
Db 1 QGLCARALDYQADADDEISFDPENLITGIEVIDEGMWRGYPDGHFGKXPANYVELI 58

## RESULT 8

US-09-925-299-1334  
Sequence 1334, Application US/09925299  
Patent No. US2002005627A1  
GENERAL INFORMATION:

APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA102

CURRENT APPLICATION NUMBER: US/09/925,299

CURRENT FILING DATE: 2001-08-10

PRIOR APPLICATION NUMBER: PCT/US00/05883

PRIOR FILING DATE: 2000-03-08

PRIOR APPLICATION NUMBER: 60/124,270

PRIOR FILING DATE: 1999-03-12

NUMBER OF SEQ ID NOS: 1556

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 1334

LENGTH: 55

TYPE: PRT

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: SITE

LOCATION: (10)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (49)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (52)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (54)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (55)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (55)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

NAME/KEY: SITE

LOCATION: (55)

OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-925-299-1334

Query Match

Best Local Similarity 6.0%; Score 241; DB 9; Length 55;  
Matches 47; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 69 CGPVTXHSYSDRFTDVRGSGPRGALASXAPARRGATRSQXAXWG 120  
Db 2 CGPVTXHSYSDRFTDVRGSGPRGALASGAPARRGATRSQXAXWG 53

## RESULT 9

US-09-925-299-1334  
Sequence 1334, Application US/09925299  
Patent No. US2002005627A1  
GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

FILE REFERENCE: PA102

CURRENT APPLICATION NUMBER: US/09/925,299

CURRENT FILING DATE: 2001-08-10

PRIOR APPLICATION NUMBER: PCT/US00/05883

PRIOR FILING DATE: 2000-03-08

PRIOR APPLICATION NUMBER: 60/124,270

PRIOR FILING DATE: 1999-03-12

NUMBER OF SEQ ID NOS: 1556  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO: 1334  
 LENGTH: 55  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: SITE  
 LOCATION: (10)  
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 NAME/KEY: SITE  
 LOCATION: (49)  
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 NAME/KEY: SITE  
 LOCATION: (52)  
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 NAME/KEY: SITE  
 LOCATION: (54)  
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 NAME/KEY: SITE  
 LOCATION: (55)  
 OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 US-09-925-299-1334

Query Match  
 Best Local Similarity 90.4%; Score 241; DB 10; Length 55;  
 Matches 47; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 69 CDPPTKHYSDREFXTPRRGEGPRGALASAKRPPARPGATSGXXARNG 120  
 DB 2 CDPPTKHYSDREFXTPRRGEGPRGALASAKRPPARPGATSGXXARNG 53

RESULT 10  
 US-09-879-957-26  
 Sequence 26, Application US/09879957  
 Patent No. US20020034755A1  
 GENERAL INFORMATION:  
 APPLICANT: SPARKS, Andrew B.  
 HOFFMAN, No. US20020034755A1h  
 KAY, Brian K.  
 FOMKES, Dana M.  
 MCCONNELL, Stephen J.  
 TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
 DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
 USING SAME  
 NUMBER OF SEQUENCES: 227  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Pennie & Edmonds LLP  
 STREET: 1155 Avenue of the Americas  
 CITY: New York  
 STATE: New York  
 COUNTRY: USA  
 ZIP: 10036-2721  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/879,957  
 FILING DATE: 13-Jun-2001  
 CLASSIFICATION: <Unknown>  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/630,915  
 FILING DATE: 03-APR-1996  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Mierock, S. Leslie  
 REGISTRATION NUMBER: 18,872  
 REFERENCE/DOCKET NUMBER: 1101-174  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (212) 790-9090  
 TELEFAX: (212) 869-8864/9741

TELEX: 66141 PENNIE  
 INFORMATION FOR SEQ ID NO: 26:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 486 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: <Unknown>  
 TOPOLOGY: unknown  
 MOLECULE TYPE: peptide  
 SEQUENCE DESCRIPTION: SEQ ID NO: 26:  
 US-09-879-957-26

Query Match  
 Best Local Similarity 29.5%; Score 228.5; DB 9; Length 486;  
 Matches 85; Conservative 24; Mismatches 122; Indels 57; Gaps 11;

QY 275 GAGG--AKXSVGLXNLNASKSESRPGXTIROR--RGASVGLGXFXRLSPPARGP--PLH 328  
 DB 234 GARGLKAKFESLAEKREBEKAEQAQMAQQQERKAVYMSREVQQPMVEEPAPAPQ 293  
 QY 329 APRKRGARAPPRAGTGVRCXVFLVGNARPERRRPKXSPRTASWGTRKTXSISFQ 368  
 DB 294 LPKTISSEWVP-----AESHLPESSPPVSRREY-----PV 325  
 QY 389 PG-KLRSPFLQXQLOPETHFGREPAAXSR--PRADLPAREPAPSPCLVQAEBAVYE 446  
 DB 326 PSLPTRQSPQNHEDNE-----EPALPRTREGLVYBEFVYEAAPFL--EPEPEPVE 379  
 QY 447 EPXQETFFRQPLVYQOQXXGSEIHDIH-----QGQGLXQGLCARAYD 492  
 DB 380 PEPETPEPVEVGLDQDADAGDYEDVLEPEDTSLSYQAGPSAGAGAGISALALYD 439  
 QY 493 YQAADXEISFDPENLITGIEVXXEGWNRGYPGDFGKXPANYVELI 540  
 DB 440 YQGESEDELSPDPDITITIDEMVDEGWRGQ--CRGHGFLPANYVELL 486

RESULT 11  
 US-10-369-493-5931  
 Sequence 5931, Application US/10369493  
 Publication No. US2003023675A1  
 GENERAL INFORMATION:  
 APPLICANT: Cao, Yongwei  
 APPLICANT: Hinkle, Gregory J.  
 APPLICANT: Slater, Steven C.  
 APPLICANT: Goldman, Barry S.  
 APPLICANT: Chen, Xiandeng  
 TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
 FILE REFERENCE: 38-10(520521)B  
 CURRENT APPLICATION NUMBER: US/10/369,493  
 CURRENT FILING DATE: 2003-02-28  
 PRIOR APPLICATION NUMBER: US 60/360,039  
 PRIOR FILING DATE: 2002-02-21  
 NUMBER OF SEQ ID NOS: 47574  
 SEQ ID NO 5931  
 LENGTH: 643  
 TYPE: PRT  
 ORGANISM: Caenorhabditis elegans  
 US-10-369-493-5931

Query Match  
 Best Local Similarity 27.5%; Score 226.5; DB 15; Length 643;  
 Matches 82; Conservative 20; Mismatches 75; Indels 121; Gaps 13;

QY 319 SPFAGRPPLHAPRXRGARAPPRAGTGVRCXVFLVGNARPERRPXK-----PS--- 369  
 DB 391 SKPSG-FVLKPKQVNGSPKMPVGT-----TSPREPVNRLTEPDEPSTYT 436  
 QY 370 -----RTASWGTRX-----TXSISQPGKLSPP--FL 397  
 DB 437 PKPAYEPBMVYKPEKMKAVSYDAYEPPAPAPPTLATPTVIAFPPEPPTAPASIV 496  
 QY 398 QXQLOPETHFGREPA--XXSRP-----RADLPA----- 425

Db 497 ASQYAPVHSEFEEVAVPVSAFSAQYDAPPEPISHSSSSQLPHAIASQYDMPV 556  
Qy 426 -EAP--AESXPPCLVOAEAEVAYEPEEQETFEQPPVVOQXXGSEHIDHHIOGGLSX 482  
Db 557 PEEVPAFAPKSPINXAA-----PPIDQ---YDFPFAVARN-----589  
Qy 483 QGLCARALYDQADDXEISDPENLTGIEVXXGKMGKRGCPDGHFGKXPANTYVELI 540  
Db 590 ----AMALWDYQADDTISFDPDIDIDQVDSGKMGKRGVGLFPANTYVELI 643

RESULT 12  
US-09-879-957-14  
Sequence 14, Application US/09879957  
Patent No. US20020034755A1  
GENERAL INFORMATION:  
APPLICANT: SPARKS, Andrew B.  
HOFFMAN, No. US20020034755A1h  
KAY, Brian K.  
FOWLES, Dana M.  
MCCONNELL, Stephen J.  
TITLE OF INVENTION: POLYPEPTIDES HAVING A FUNCTIONAL  
DOMAIN OF INTEREST AND METHODS OF IDENTIFYING AND  
USING SAME  
NUMBER OF SEQUENCES: 227  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Penile & Edmonds LLP  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: USA  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/879,957  
FILING DATE: 13-Jun-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/630,915  
FILING DATE: 03-APR-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Mierock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 1101-174  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 790-9090  
TELEFAX: (212) 869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 546 amino acids  
TYPE: amino acid  
STRANDEDNESS: <Unknown>  
MOLECULE TYPE: peptide  
TOPOLOGY: unknown  
SEQUENCE DESCRIPTION: SEQ ID NO: 14:  
US-09-879-957-14

Query Match 5.2%; Score 208; DB 9; Length 546;  
Best Local Similarity 35.6%; Pred. No. 1.3e-09;  
Matches 57; Conservative 18; Mismatches 57; Indels 28; Gaps 6;  
Qy 386 SPQPKLSPFLQXQLOTPETHFGREPAAXXSRPADLPASEPAPSPPCVQAEAEVY 445  
Db 407 SQQPIEDPFP--SSIVYDAPAFKAPSPYRGSEF-----EFYS-----IEAGIP 450  
Qy 446 EEPXQETFEQPPVVOQXXGSEHIDHHIOGGLSX-----ISXQGLCARALYDQADDX 499

Db 451 EAGSQGLTYTSEPVTE-----TTEAPGHVQAEEDYIDGESPLGITAIALYDQAGDD 505  
Qy 500 EISFDPENLTGIEVXXGKMGKRGCPDGHFGKXPANTYVELI 539  
Db 506 EISFDPDITITNEMIDGGMWGRV-CKGRYGLFPANTYVELI 544

RESULT 13  
US-09-925-299-1277  
Sequence 1277, Application US/09925299  
Patent No. US20020055627A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA102  
CURRENT APPLICATION NUMBER: US/09/925,299  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05883  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1556  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1277  
LENGTH: 40  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (33)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-299-1277

Query Match 4.8%; Score 192; DB 9; Length 40;  
Best Local Similarity 90.0%; Pred. No. 1.7e-09;  
Matches 36; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 234 WDXTVVROVSFTLLMKCCCHGNPAQYERRRRXXHLYTVLG 273  
Db 1 WYTVVROVSFTLLMKCCCHGNPAQYERRRRXXHLYTVLG 40

RESULT 14  
US-09-925-299-1277  
Sequence 1277, Application US/09925299  
Patent No. US20030040617A9  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA102  
CURRENT APPLICATION NUMBER: US/09/925,299  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05883  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1556  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1277  
LENGTH: 40  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (33)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-299-1277

Query Match 4.8%; Score 192; DB 10; Length 40;  
Best Local Similarity 90.0%; Pred. No. 1.7e-09;  
Matches 36; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
Qy 234 WDXTVVROVSFTLLMKCCCHGNPAQYERRRRXXHLYTVLG 273

Db 1 WYTTVROVSTLLMCCCHGNPAQYERRRFXHLVYVLG 40

RESULT 15  
US-10-369-493-4212  
; Sequence 4212, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 4212  
; LENGTH: 769  
; TYPE: PRT  
; ORGANISM: Neurospora crassa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(769)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-4212

Query Match 4.8%; Score 190; DB 15; Length 769;  
Best Local Similarity 24.9%; Pred. No. 7,4e-08;  
Matches 97; Conservative 38; Mismatches 164; Indels 90; Gaps 19;

QY 314 PXXRLSPGPRPLHPRXGRAP-----PRAGTVRCXVPVFLGNGA 358  
Db 414 PEAPASBAG--ISALREFKDTAFAVARTGSRASPSVSPGPAATGECHTPPLPTGS 470  
QY 359 RPE--RRPYKRPSTARSMTWXTXSISPPQPK-LSPFLQXQLTQPTHFGRPA- 413  
Db 471 RPSGGFALPGL-PSRP-----SVADDEGNDEROLEEVRALPEPADQHEEPSS 518  
QY 414 -AXXSRL-PAULPAEPPASXPCUQAEEAVEE--PXQETFEQPLVQOQXGS 468  
Db 519 PARIAVEIASAVPEVEPPAGNLPPRPVPSQDLPKESDLPVEEDT--HDPRAALNVA 575  
QY 469 EHIDHITQGGLSXQ-----GLCARLYDQAADXEISFPENLITGIEVXXEGWVG 522  
Db 576 ESL-----GGQVEVQAGSDGKRALVQYDEKADNEIDLQEGDYVTNIQWVDDMMWG 630  
QY 523 YGPDGHTGKAPANTVELIEKGXGHIILFSPQTLFPCWKRKPKXHSALFQEXDPQXGX 582  
Db 631 TNAQGESGLPSSNVEVEDEDEPAAPSAPT--VAATQAFVPSAQAQAV----- 680  
QY 583 LRAPSGLADSACHFCKX-----NG--LVIPTPSCIRPXX-----TAWLLPLT 624  
Db 681 --APATPADQVAGTATAQFDYBAEDNGKXCLEAPHISAAXRPSSDXRIGTLSFP-E 737  
QY 625 GYXAKPCXWSPFWPLPSCGEGSXGASG 653  
Db 738 GATITGDEFDDWVF-----GHYSASG 761

Search completed: April 27, 2004, 11:05:13  
Job time : 41.7311 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

CM protein - protein search, using sw model

Run on: April 27, 2004, 10:46:49 ; Search time 17.097 Seconds

(without alignments)  
4461.589 Million cell updates/sec

Title: US-10-028-952a-9

Perfect score: 3989

Sequence: 1 HEIPVPTVYPAKQXXERA.....DFRKKKKKKKKLEKWTSSX 793

Scoring table:

BLOSUM62

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database:

1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	234	5.9	486	2	hematopoietic line
2	228.5	5.7	486	2	LckBPI protein - m
3	226.5	5.7	643	2	hypothetical prote
4	208	5.2	546	2	coractin - mouse
5	204.5	5.1	563	2	src substrate p80/
6	202	5.1	550	2	mammary tumor/sgna
7	151.5	3.8	1111	1	myosin heavy chain
8	148.5	3.7	390	2	Abi binding protei
9	143	3.6	315	2	35k proline-rich p
10	143	3.6	592	1	actin-binding prot
11	141	3.5	290	2	hypothetical prote
12	134.5	3.4	617	2	actin-binding prot
13	132	3.3	261	2	nebulin, skeletal
14	131	3.3	705	2	synapsin I eplice
15	126	3.2	6669	2	synapsin Ia - h
16	123	3.1	1181	2	myosin IC - slime
17	121.5	3.0	1107	1	myosin I heavy cha
18	120	3.0	670	2	probable signal tr
19	117.5	2.9	540	2	signal transducing
20	117.5	2.9	704	2	synapsin Ia - rat
21	116	2.9	1270	2	adaptor protein in
22	115.5	2.9	1290	2	1-phosphatidylinos
23	115.5	2.9	1291	1	1-phosphatidylinos
24	114.5	2.9	1168	1	myosin heavy chain
25	114	2.9	339	2	epidermal growth f
26	114	2.9	158394	2	c-Crk - mouse
27	114	2.9	1109	2	myosin-IC (similar
28	113.5	2.8	330	2	Grb-2 related adap
29	113.5	2.8	452	2	hypothetical prote

30	113.5	2.8	1290	2	A36465	1-phosphatidylinos
31	112.5	2.8	548	2	UC4917	signal transducing
32	112	2.8	304	2	A45022	CRK-II - human
33	112	2.8	1099	2	A59300	myosin-Ic - mouse
34	111.5	2.8	269	2	UC4899	proline rich prote
35	111.5	2.8	1099	1	UC4899	myosin IB heavy ch
36	111.5	2.8	1215	2	T32734	myosin-1A - Acanth
37	110.5	2.8	825	2	T23612	hypothetical prote
38	110	2.8	940	2	T00056	hypothetical prote
39	110	2.8	1006	2	T00050	hypothetical prote
40	109	2.7	1460	1	ED881F	immediate-early pr
41	108.5	2.7	367	2	S50621	SSUB1 protein - ye
42	108.5	2.7	629	2	T14776	hypothetical prote
43	107.5	2.7	209	2	S44650	f42n10.3 protein -
44	106	2.7	303	2	S58952	SH2/SH3 adaptor pr
45	105	2.6	379	2	S31719	proline-rich prote

#### ALIGNMENTS

##### RESULT 1

S07633 hematopoietic lineage cell-specific protein Hs1 - human  
C/Species: Homo sapiens (man)  
C/Date: 29-Jan-1993 #sequence\_revision 29-Jan-1993 #text\_change 05-Nov-1999  
C/Accession: S07633; A47478; B47478; C47478; D47478; E47478  
R/Kitamura, D.; Kaneko, H.; Miyagoe, Y.; Ariyasu, T.; Watanabe, T.  
Nucleic Acids Res. 17, 9367-9379, 1989  
A/Title: Isolation and characterization of a novel human gene expressed specifically in  
A/Reference number: S07633; MIMD:90067924; PMID:2587259  
A/Accession: S07633  
A/Molecule type: mRNA  
A/Residues: 1-486 <KIT>  
A/Cross-references: EMBL:X1663; NID:932054; PIDN:CA34651.1; PID:932055  
R/Yamanashi, Y.; Okada, M.; Samba, T.; Yamori, T.; Umemori, H.; Tsunawawa, S.; Toyoshita  
Proc. Natl. Acad. Sci. U.S.A. 90, 3631-3635, 1993  
A/Title: Identification of Hs1 protein as a major substrate of protein-tyrosine kinase  
A/Reference number: A47478; MIMD:93234551; PMID:7682714  
A/Accession: A47478  
A/Status: preliminary  
A/Molecule type: protein  
A/Residues: 4-19, 'XXX', '23-26', 'X', '79-93', 'X', '95', 'X', '134-146', '208-223', '274-284', 'X', '286', 'X', '2  
A/Experimental source: Daudi, B-lymphoblastoid cells  
A/Note: sequence modified after extraction from NCBI backbone  
C/Superfamily: SH3 homology  
F/435-482/Domain: SH3 homology <SH3>

Query Match 5.9%; Score 234; DB 2; Length 486;

Best local similarity 26.8%; Pred. No. 1.8e-12;

Matches 84; Conservative 25; Mismatches 97; Indels 108; Gaps 12;

QY	275	GANG--AKLXSVGLXNASKSESRFGKTRORRGASVGLGKPKXKSPRAGR-----	325
DB	234	GARGLAKKESVNAEKKREKEEKAQVARROQERRA-----VTKSSPAPQVIMSEEP	288
QY	326	--PLHAPKXGRARAPRAGTGVRCVAPFVLGNGARPERPXXRPRTARSWGTWXTXS	383
DB	289	AVAPAPPKTKSSAWP-----VGTPTSE--SEPTATREHHV-----	325
QY	384	ISSPQPKLRSPLQXOLTOPETHFGREPAAXXSRRADLPABEPAPSPAPCL--VOAE	440
DB	326	-----PLLPTRQTLFEDN-----EEP--PALPPTLDELQVE	355
QY	441	EEAVVEEPKQETFFQYQPLVQOQXXGSEHIDHIQ-----	476
DB	356	EEVVEEAEPPPP--EPPEPPNDYDEVEMDNHEEDPEEGVEEVLPEEDSFFSALA	413
QY	477	-----GGGLXQGLCARALDYQADDXISFDPENLTGIVXXEGWRRGGPGG	527
DB	414	GSSGCPAGAGAGAVALLGISVALLDYQGGSDLSFDPDVITDIMEVDEGWRG--RCGG	472
QY	528	HFGMPKPNVYELLE	541

Db 473 HFGLPFANYVXKLE 486

## RESULT 2

LckBp1 protein - mouse  
 N:Alternate names: HSI  
 C:Species: Mus musculus (house mouse)  
 C>Date: 29-May-1998 #sequence\_revision 29-May-1998 #text\_change 05-Nov-1999  
 C:Accession: I49760; S57244; S52474  
 R:Kikemura, D.; Kaneko, H.; Taniguchi, I.; Akagi, K.; Yamamura, K.; Watanabe, T.  
 B:Biochem. Biophys. Res. Commun. 208, 1137-1146, 1995  
 A>Title: Molecular cloning and characterization of mouse HSI.  
 A:Reference number: I49760; MUID:95217198; PMID:7535527  
 A:Accession: I49760  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-486 <RES>  
 A:Cross-references: GB:D42120; NID:9806521; PIDN:BA0701.1; PID:9806522  
 R:Takemoto, Y.; Furuta, M.; Li, X.K.; Strong-Sparks, W.O.; Hashimoto, Y.  
 EMOB J. 14, 3403-3414, 1995  
 A>Title: LckBp1, a proline-rich protein expressed in haematopoietic lineage cells, direct  
 A:Reference number: S57244; MUID:95354658; PMID:7628441  
 A:Accession: S57244  
 A>Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-486 <RES>  
 A:Cross-references: EMBL:X84797; NID:9683480; PIDN:CA59265.1; PID:9683481  
 A:Superfamily: SH3 homology  
 F:436-483/Domain: SH3 homology <SH3>

Query Match 5.7%; Score 228.5; DB 2; Length 486;  
 Best Local Similarity 29.5%; Pred. No. 5.3e-12;  
 Matches 85; Conservative 24; Mismatches 122; Indels 57; Gaps 11;  
 275 GANG-AKLXSVGLINAKSESERPGXITRQR--RCASVGLGXPPXRLSPPAGRP--PLH 328  
 234 GAGGLAKXEPFLAEKREKKEEKAQOMARQOCERKAVVMKREVRVQSPMPEEPAAPQ 293  
 QY 329 APXKRGARAPPAGGCVFVGLGNGARPPRRRXXKPSRTARWGTWRTXISSPQ 388  
 Db 294 LPKXISSEWVP-----AESHLPESSQPVSSRREY-----PV 325  
 QY 389 PG-KLRSPFQXQLOTPETHFGREPAAXSR--PRADLPREEPAAXPCLVQAEBAVYE 446  
 Db 326 PSAPTQSPLOQNHLEDN-----EPFALPRTTEGLQVVEPYEAAPL--EPEPEPVE 379  
 QY 447 EPEQETFEYQPPVVOQXXGSEHIDHHT-----QGGGLSXQGLCARALYD 492  
 Db 380 PEPETPEPDYEDVGLRQDADAGDYEDVLEPEPTSLSYQAGPSAGAGISALALYD 439  
 QY 493 YQADADKEISFPDENLITGIEVXXEGMWRGYPGDGHFGKXKPANYVELI 540  
 Db 440 YQGGSELSFDPDDITTDIEMVDEGMWRQ--CRGHGFLFPANYVELL 486

## RESULT 3

T23453  
 hypothetical protein K08E3.4 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
 C:Accession: T23453  
 R:McMurray, A.  
 submitted to the EMBL Data Library, November 1996  
 A:Reference number: Z19743  
 A:Accession: T23453  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-643 <ML>  
 A:Cross-references: EMBL:Z81566; PIDN:CA804592.1; GSPDB:GN00021; CESP:K08E3.4  
 C:Experimental source: clone K08E3  
 C:Genetics:

A:Gene: CESP:K08E3.4  
 A:Map position: 3  
 A:Introns: 47/1; 84/3; 185/2; 255/2; 455/1

Query Match 5.7%; Score 226.5; DB 2; Length 643;  
 Best Local Similarity 27.5%; Pred. No. 1.1e-11;  
 Matches 82; Conservative 20; Mismatches 75; Indels 121; Gaps 13;

QY 319 SPAPGPPPLHAPKXRGARAPPAGGCVFVGLGNGARPPRRRXXR-----PS----- 369  
 Db 391 SKPSG-FVLFKPPQVNSPKXKMPVGT-----TSPREVNLTPEDEPSTYT 436  
 QY 370 -----RTARSGTWRX-----TXSISPPQGLNSP--EL 397  
 Db 437 PKAIAPPEPMVYKPEAMKPAVSQYDAVEBPAPAPPTLAPPVLAPEPPTAPSHY 496  
 QY 398 QXQLOTPETHFGREPA--XXSRP-----RADLPA----- 425  
 Db 497 ASQYDAPPVHESEFEPVPPVPSAPSHVASCYDAPPEDIDSHSSSQLPALIASQYDMPVYM 556  
 QY 426 -EPP--APSKPCLVQAEBAVYEEPEQETFEYQPPVVOQXXGSEHIDHHTIQGGLSX 482  
 Db 557 PEEVFPAPKSSPFLKVA-----PPIQ-----YDFPVAEQN----- 589  
 QY 483 QGLCARALYDQADDXEISFPDENLITGIEVXXEGMWRGYPGDGHFGKXKPANYVELI 540  
 Db 590 -----AMALMDYQADDTIEISFDDITTDIDQVDSGMWKGKRAPSGVGLFPANYVELI 643

## RESULT 4

148899  
 cortactin - mouse  
 C:Species: Mus musculus (house mouse)  
 C>Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 05-Nov-1999  
 C:Accession: I48899; A49514  
 R:Milarese, M.R.; Mamion-Henderson, J.; Wu, H.; Parsons, J.T.; Bender, T.P.  
 Oncogene 9, 1869-197, 1994  
 A>Title: The protein tyrosine kinase substrate cortactin is differentially expressed in  
 A:Reference number: I48899; MUID:94268839; PMID:7516062  
 A:Accession: I48899  
 A>Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-546 <RES>  
 A:Cross-references: EMBL:U03184; NID:9414990; PIDN:AA19669.1; PID:9509495  
 R:Zhan, X.; Hu, X.; Hampton, B.; Burgess, W.H.; Friesel, R.; Maciag, T.  
 J. Biol. Chem. 268, 24427-24431, 1993  
 A>Title: Murine cortactin is phosphorylated in response to fibroblast growth factor-1  
 A:Reference number: A49514; MUID:94043284; PMID:7693700  
 A:Accession: A49514  
 A>Status: preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid; protein  
 A:Residues: 1-8, 'R', 10-297, 'S', 299-546 <ZHA>  
 A:Experimental source: BALB/c 3T3 cells  
 A>Note: sequence extracted from NCBI backbone (NCBIP:139492)  
 C:Superfamily: SH3 homology  
 F:495-542/Domain: SH3 homology <SH3>

Query Match 5.2%; Score 208; DB 2; Length 546;  
 Best Local Similarity 35.6%; Pred. No. 3.6e-10;  
 Matches 57; Conservative 18; Mismatches 57; Indels 28; Gaps 6;

QY 386 SPQFKLSPLQLOTPETHFGREPAAXSRPRADLPREEPAAXPCLVQAEBAVY 445  
 Db 407 SPQPIEDRP--SSPIYEDAPFAEBSYRSGE-----EPEYS-----TEAGIP 450  
 QY 446 EPEQETFEYQPPVVOQXXGSEHIDHHTIQGG-----LSXQGLCARALYDQADDX 499  
 Db 451 EAGSQGGLTYSEPYE-----TEAPGHTQABDDTQIGESDGLTALALYDQAGDD 505  
 QY 500 EISFPDENLITGIEVXXEGMWRGYPGDGHFGKXKPANYVEL 539  
 Db 506 EISFPDDIINIMIDGMWRG--CKGRVGLFPANYVEL 544

## RESULT 5

src substrate p80/85 - chicken  
N:Alternate names: kinase-related transforming protein substrate  
C:Species: Gallus gallus (chicken)  
C>Date: 03-Aug-1992 #sequence\_revision 03-Aug-1992 #text\_change 21-Jul-2000  
R:Accession: A41530  
R:Rev. H.; Reynolds, A.B.; Kanner, S.B.; Vines, R.R.; Parsons, J.T.  
Mol. Cell. Biol. 11, 5113-5124, 1991  
A:Title: Identification and characterization of a novel cytoskeleton-associated pp60(src  
A:Reference number: A41530; MUID:92017786; PMID:1922035  
A:Accession: A41530  
A:Molecule type: mRNA  
A:Residues: 1-563 <MID>  
A:Cross-references: GB:M73705; NID:g212588; PIDN:AAA49031.1; PID:g212589  
C:Superfamily: SH3 homology  
C:Keywords: phosphoprotein; tandem repeat  
F:92-128/Domain: src substrate repeat <SR1>  
F:129-165/Domain: src substrate repeat <SR2>  
F:166-202/Domain: src substrate repeat <SR3>  
F:203-239/Domain: src substrate repeat <SR4>  
F:240-276/Domain: src substrate repeat <SR5>  
F:277-313/Domain: src substrate repeat <SR6>  
F:314-334/Domain: src substrate repeat #status atypical <SR7>  
F:512-559/Domain: SH3 homology <SH3>

Query Match 5.1%; Score 204.5; DB 2; Length 563;  
Best Local Similarity 36.1%; Pred. No. 7.4e-10;  
Matches 52; Conservative 18; Mismatches 45; Indels 29; Gaps 5;

QY 424 PAEEPPAPXP--PCLVQAEBAVEEPEX-----EGETFE-----QPLVQOQXXGS 468  
DB 419 PAEKPTESSPYQDAVSADASAYKNSSTYSABHEBESGKTTGSDVQSAVSQREAYE 478  
QY 469 EHIHNIQOQGLSKQ-----GLCARLYVQADDXEISPDENITIGLEYX 515  
DB 479 PETVEVAGADHYQAEENTYDEYENELGITALVYQAGDDEISFDDITITENT 538  
QY 516 XEGWNRGYGPDGHFGKXPANYVEL 539  
DB 539 DDGWMRGV-CKGRYGLFPANYVEL 561

## RESULT 6

mammary tumor/squamous cell carcinoma-associated protein EMS1 - human  
A:Accession: A48063  
C:Species: Homo sapiens (man)  
C>Date: 24-Feb-1994 #sequence\_revision 18-Nov-1994 #text\_change 29-Aug-1997  
R:Accession: A48063  
R:Schuring, E.; Verhoeven, E.; Litvinov, S.; Michalides, R.J.  
Mol. Cell. Biol. 13, 2891-2898, 1993  
A:Title: The product of the EMS1 gene, amplified and overexpressed in human carcinomas,  
A:Reference number: A48063; MUID:93233650; PMID:8474448  
A:Accession: A48063  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-550 <SCH>  
A>Note: Sequence extracted from NCBI backbone (NCBIP:129872)  
A:Genetics:  
A:Gene: GDB:EMS1  
A:Cross-references: GDB:202962; OMIM:164765  
A:Map position: 11q13-11q13  
C:Superfamily: SH3 homology  
F:499-546/Domain: SH3 homology <SH3>

Query Match 5.1%; Score 202; DB 2; Length 550;  
Best Local Similarity 36.0%; Pred. No. 1.2e-09;  
Matches 58; Conservative 17; Mismatches 60; Indels 26; Gaps 6;  
QY 386 SPOGKLRSPFLQQLQTPETHFGREPAAXXSRRPADLPAEPPAPXPCLVQAEBAVY 445  
DB 407 APQPTREHLP-----SSPV-----YEDASF---KALSTYKGPVSGTPEPVYGMENADY 453

QY 446 EEPXED-----ETPEPPVLOQXXGSEHIDHIOGGSXQGLCARLYVQADDD 498  
DB 454 REASSQOGLAVTEAYE-----SAEPGHYPAEOSTYDEYENDIGYTRVALYDQAGD 508  
QY 499 XEISPDENITIGLEYXEGWNRGYGPDGHFGKXPANYVEL 539  
DB 509 DEISFPDDITINIMIDGWMRGV-CKGRYGLFPANYVEL 548

## RESULT 7

A33284  
myosin heavy chain IB - slime mold (Dictyostelium discoideum)  
N:Conting: myosin ATPase (EC 3.6.4.1)  
C:Species: Dictyostelium discoideum  
C>Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 19-Apr-2002  
R:Accession: A33284  
R:Jung, G.; Saxe III, C.L.; Kimmel, A.R.; Hammer III, J.A.  
Proc. Natl. Acad. Sci. U.S.A. 86, 6186-6190, 1989  
A:Title: Dictyostelium discoideum contains a gene encoding a myosin I heavy chain.  
A:Reference number: A33284; MUID:89345628; PMID:2762320  
A:Accession: A33284  
A:Molecule type: DNA  
A:Residues: 1-1111 <JUN>  
A:Cross-references: GB:M26037; NID:g167838; PIDN:AAA33229.1; PID:g167839  
C:Genetics:  
A:Insertions: 1/3; 39/3  
C:Superfamily: protozoan myosin heavy chain IB; myosin motor domain homology; SH3 homol  
C:Keywords: actin binding; ATP; hydrolase; nucleotide binding; P-loop; phosphoprotein;  
F:12-678/Domain: myosin motor domain homology <MOT>  
F:102-109/Region: nucleotide-binding motif A (P-loop)  
F:566-587/Region: actin binding #status predicted  
F:712-1111/Domain: carboxyl-terminal <CTD>  
F:712-901/Region: basic  
F:922-1058/Region: alanine/glutamine/glycine/proline-rich  
F:1060-1108/Domain: SH3 homology <SH3>  
F:108/Binding site: ATP (Lys) #status predicted  
F:332/Binding site: phosphate (Ser) (covalent) #status predicted

Query Match 3.8%; Score 151.5; DB 1; Length 1111;  
Best Local Similarity 21.7%; Pred. No. 5.9e-05;  
Matches 57; Conservative 21; Mismatches 86; Indels 99; Gaps 5;

QY 291 SKSESPPGXTIRQBGASVGLXPXXRLSPRPPLVAPRGARAPPRAGTGRGV 350  
DB 931 SQSSKPA---PQSAGKRGMPQG---AGQPPQPPQOQCPMPQPPQGGARXP 981  
QY 351 PVLNGARPERRPXXRPRTARSWGTXYSISSPPGKLRSPFLQQLQTPETHFR 410  
DB 982 PQQGGGARP-----MGAPQGGAPQOQAGRQLPPTQGGGA 1018  
QY 411 E-----PAAXSRPADLPAEPPAPXPCLVQAEBAVVEEPEXQETFE 455  
DB 1019 PGGGAPWGRGAPGGGPPAGGRPLPYAKRPQSRP----- 1056  
QY 456 EQPPLVQOQXXGSEHIDHIOGGSXQGLCARLYVQADDXEISPDENITIGLEYX 515  
DB 1057 -----TAKALYDYDASSTDELSEFGEDIIIFVQGD 1086  
QY 516 XEGWNRGYGPDGHFGKXPANYVE 538  
DB 1087 NGQMTQGLKSGQKGMWPTNYDQ 1109

## RESULT 8

G01936  
Abi binding protein 3 - human  
C:Species: Homo sapiens (man)  
C>Date: 21-Dec-1996 #sequence\_revision 06-Jun-1997 #text\_change 17-Jul-1998  
C:Accession: G01936  
R:Ren, R.  
submitted to the EMBL Data Library, July 1995  
A:Reference number: G08875



C:Species: Schizosaccharomyces pombe  
 C:Date: 11-Jan-2000 #sequence\_revision 11-Jan-2000 #text\_change 21-Jul-2000  
 C:Accession: T42526  
 R:Yoshioaka, S.; Kato, K.; Nakai, K.; Okayama, H.; Nojima, H.  
 DNA Res. 4, 363-369, 1997  
 A:Title: Identification of open reading frames in Schizosaccharomyces pombe cDNAs.  
 A:Reference number: Z17323, PMID:96162722, PMID:9501991  
 A:Accession: T42526  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-290 <YOS>  
 A:Cross-references: EMBL:D89164; NID:g1749535; PIDN:BA13826.1; PID:g1749536  
 A:Experimental source: strain PR745

Query Match 3.5%; Score 141; DB 2; Length 290;  
 Best Local Similarity 27.7%; Pred. No. 0.00012;  
 Matches 52; Conservative 25; Mismatches 59; Indels 52; Gaps 11;

QY 387 POP-GKLRSPFLQXQLTQPEHFGREPAAXXSRRPADPA-EEDAPSPCLVQAEEDAV 444  
 DB 121 PEPAQQLNEPVVPPPPHDET---QEPVVG-----GDVKAETHTQPTKTPAIV-----I 166  
 QY 445 YEEPYEOTFEYQEPPLVQ-QQXXGSEHID-----HHIQ-QGQ----- 480  
 DB 167 YDYSPEENEIE---LVENEQLQLEFVDGWLGENKGGQGLFPPSNVYEITGNETAN 223  
 QY 461 -----SXQGLCARALYDQADDXEISFDPELTLGIEVXXEGWGRGYPGPHGKMX 532  
 DB 224 NPPAPQAGGPGKSKYKALDYQAEQDNELSFEDETILANVDCVDPNMEG-ECHGRGLF 282  
 QY 533 PANYVELI 540  
 DB 283 PSNYVEEI 290

## RESULT 12

S42719  
 actin-binding protein ABPI - Saccharomyces exiguus  
 C:Species: Saccharomyces exiguus  
 C:Date: 07-Sep-1999 #sequence\_revision 19-Apr-1996 #text\_change 13-Aug-1999  
 C:Accession: S42719  
 R:Langue, U.; Steiner, S.; Grolig, F.; Wagner, G.; Philippen, P.  
 Biochim. Biophys. Acta 1217, 214-218, 1994  
 A:Title: Cloning and sequencing of a gene coding for an actin binding protein of Sacchar.  
 A:Reference number: S42719; PMID:94154001; PMID:8110838  
 A:Accession: S42719  
 A:Molecule type: DNA  
 A:Residues: 1-617 <LAN>  
 A:Cross-references: EMBL:X73977; NID:g433513; PIDN:CAA52156.1; PID:g433514  
 C:Superfamily: actin-binding protein ABPI; SH3 Homology  
 C:Keywords: actin binding; cytoskeleton  
 P:564-613/Domain: SH3 homology <SH3>

Query Match 3.4%; Score 134.5; DB 2; Length 617;  
 Best Local Similarity 25.0%; Pred. No. 0.00095;  
 Matches 52; Conservative 21; Mismatches 82; Indels 53; Gaps 8;

QY 385 SSPQPKLRSPFLQXQLTQ-PETH-----FGREPAAXXSRRPADIPAE 426  
 DB 408 SKPATQEVKKPPTPENIGQLPQKMTETPEHEEDNDNDGDEDEDEPPKRNIPPVMPAR 467  
 QY 427 EAPSP--XPCLVQAE--EAVYEEFXQETFYEP-----PLVQQXXGSEHID 472  
 DB 466 ESAPQQLPPEPNTPEPVEEGSEEEEBEESAPASLPSRNAAPPEPEQOESEEE 527  
 QY 472 HHIOGGLSXQGLC-----ARALYDQADDXEISFDPELTLGI 512  
 DB 528 ESAPASLPSRGSVPPEPQRAVEPEBPAAEAPWATAEDYDAGDNELTLTAENDKIINI 587  
 QY 513 EVXEGWGRG-YGPHGGRKMPANYVEL 539  
 DB 588 EFVDDMWLGELETTGQKGLFSPSNVVL 615

## RESULT 13

S68234  
 Lasp-1 protein - human  
 N:Alternate names: LIM and SH3 protein  
 C:Species: Homo sapiens (man)  
 C:Date: 15-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 16-Jul-1999  
 C:Accession: S68234, S72556  
 R:Tomasetto, C.; Moog-Lutz, C.; Regnier, C.H.; Schreiber, V.; Basset, P.; Rio, M.C.  
 FEBS Lett. 373, 245-249, 1995  
 A:Title: Lasp-1 (MLN 50) defines a new LIM protein subfamily characterized by the assoc  
 A:Reference number: S68234; PMID:96033982; PMID:7589475  
 A:Accession: S68234  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-261 <TOM>  
 A:Cross-references: EMBL:X82456  
 R:Tomasetto, C.  
 submitted to the EMBL Data Library, November 1994  
 A:Reference number: S72556  
 A:Accession: S72556  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-84, 'C', 86-90, 'G', 92-219, 'D', 221-261 <T02>  
 A:Cross-references: EMBL:X82456; NID:g951273; PID:g951274  
 C:Superfamily: LIM metal-binding repeat homology; SH3 homology  
 P:5-56/Domain: LIM metal-binding repeat homology <LIM>  
 P:209-258/Domain: SH3 homology <SH3>

Query Match 3.3%; Score 132; DB 2; Length 261;  
 Best Local Similarity 29.0%; Pred. No. 0.00064;  
 Matches 42; Conservative 17; Mismatches 48; Indels 38; Gaps 7;

QY 409 GREPAAXSRP-----RADIPAEPP--APSPCLVQAEEDAV-----YEEPYEOTFEY 456  
 DB 138 GWEPERRDSDGSGSYRPLFQQQPHHTPSAPYQPPQQAQSYGKYKRA----- 190  
 QY 457 QPELVQXXGSEHIDHHIQQGLSXQGLCARALYDQADDXEISFDPELTLGIEVXX 516  
 DB 191 APVSIQRSARFG-----GGRYRAYVDYSAADDEVSFGDGTIVVQQLD 236  
 QY 517 EGWGRGYP---DGHFGKXPANYVE 538  
 DB 237 DGM--YGVERTGDTKLPDNIYE 259

## RESULT 14

A35363  
 synapsin I splice form a - human  
 C:Species: Homo sapiens (man)  
 C:Date: 17-Aug-1990 #sequence\_revision 17-Aug-1990 #text\_change 01-Dec-2000  
 C:Accession: A35363; B35363; A35805  
 R:Suedhof, T.C.  
 J. Biol. Chem. 265, 7849-7852, 1990  
 A:Title: The structure of the human synapsin I gene and protein.  
 A:Reference number: A35363; PMID:90243651; PMID:2110562  
 A:Accession: A35363

A:Molecule type: DNA  
 A:Residues: 1-705 <SUE>  
 A:Cross-references: GB:W58371; GB:J05431  
 A:Accession: B35363  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-659, 'KASPRQAP', <SU2>  
 A:Cross-references: GB:W58378; GB:J05431  
 R:Suedhof, T.C.; Hosche, C.; Oswald, R.; Kilian, M.W.  
 J. Biol. Chem. 265, 14932-14937, 1990  
 A:Title: The 5'-flanking region of the synapsin I gene. A G-C-rich, TATA- and CAAT-less  
 A:Reference number: A35805; PMID:9038667; PMID:2118519  
 A:Accession: A35805  
 A:Molecule type: DNA  
 A:Residues: 1-125 <SAU>  
 A:Cross-references: GB:W53501; NID:g338655; PIDN:AAA60608.1; PID:g553654; GB:J05630

C/Genetics:  
 A/Gene: GDB:SYN1  
 A/Cross-references: GDB:119606; OMIM:313440  
 A/Map position: Xp11.23-Xp11.23  
 C/Keywords: actin binding; alternative splicing; phosphoprotein

Query Match 3.3%; Score 131; DB 2; Length 705;  
 Best Local Similarity 24.7%; Pred. No. 0.0022;  
 Matches 76; Conservative 22; Mismatches 126; Indels 84; Gaps 13;

QY 209 VGSSTYXXAEFTKRWVHPFIGNXSWDXTVVRQVSFTLMKCCGHPAQYERNRXHLL 268  
 DB 388 VGSSTYXXAEFTKRWVHPFIGNXSWDXTVVRQVSFTLMKCCGHPAQYERNRXHLL 428  
 QY 269 VYVLXGANGAKLXSVGLKLNKSKSRSRPGXTTRORRGSVGLKXP----- 314  
 DB 429 ---GRGSHQRTSPGALPLGRQTSQQPAGPPAQCPPPQGGPPPGPGPGRQGPPLQQR 484  
 QY 315 -----XXRLSPRAGR-----PLHAPRXGRARAPPAGTGVRCXVFPVLGNGAR 359  
 DB 485 PPGQGOHLISGLPPAGSPPLPQRLPSPFTSAPQAPPAQAPPPQGGQ-RGRREVAGGPGAP 543  
 QY 360 PERRPXXRPSRTARSNGTWXT--XSISSPQPKLRS--PLQXQUTQPETHFG-----R 410  
 DB 544 PAARPPASPS-PORCAGPPQATRTQTSVSGPAPKASGAPPGQQRGPPQKPPGAPPTR 602  
 QY 411 EPAAXXSRRPADLP-AEPPASXP-----PCLVQAEAEAYEEPEQETFEQPPPLVQ 462  
 DB 603 QASQAGVPRTGPTTQQPAPSGPGPAGAPKPOLAQKPSQDV-----PPPAT 649  
 QY 463 QQXXGSEH 470  
 DB 650 AAAGPPH 657

## RESULT 15

S55024  
 nebulin, skeletal muscle - human  
 C/Species: Homo sapiens (man)  
 C/Date: 15-Jul-1995 #sequence revision 01-Sep-1995 #text\_change 05-Nov-1999  
 C/Accession: S55024; S55025; A40847  
 R/Label: S. the EMBL Data Library, January 1995  
 A/Reference number: S55024  
 A/Accession: S55024

A/Molecule type: mRNA  
 A/Residues: 1-6669 <LAB1>  
 A/Cross-references: EMBL:X83957; NID:9806561; PIDN:CA58788.1; PID:9806562  
 R/Label: S. Kolmerer, B.

U. Mol. Biol. 248, 308-315, 1995  
 A/Title: The complete primary structure of human nebulin and its correlation to muscle  
 A/Reference number: S55025; MUID:9525391; PMID:7739042  
 A/Accession: S55025

A/Status: nucleic acid sequence not shown  
 A/Molecule type: mRNA  
 A/Residues: 3731-3930-6501-6669 <LAB2>  
 A/Cross-references: EMBL:X83957

R/Label: S. Wang, K.  
 U. Biol. Chem. 266, 21215-21223, 1991  
 A/Title: Cloning, expression, and protein interaction of human nebulin fragments compose  
 A/Reference number: A40847; MUID:92042000; PMID:1682316  
 A/Accession: A40847

A/Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra  
 A/Molecule type: mRNA  
 A/Residues: 6263-6271, 'W', 6273-6329 <LIN>  
 C/Genetics:

A/Gene: GDB:NEB  
 A/Cross-references: GDB:120224; OMIM:161650  
 A/Map position: 2q24.1-2q24.2  
 C/Superfamily: human skeletal muscle nebulin; SH3 homology  
 C/Keywords: alternative splicing; skeletal muscle; tandem repeat  
 F/78-6531/Region: 35-residue repeats  
 F/6617-6666/Domain: SH3 homology <SH3>

Query Match 3.2%; Score 126; DB 2; Length 6669;  
 Best Local Similarity 27.0%; Pred. No. 0.061;  
 Matches 47; Conservative 13; Mismatches 72; Indels 42; Gaps 5;

QY 402 TOPETHFGREPAXXSRRP-----ADLPABEPASXPCLVQAE----- 442  
 DB 6503 TNPGSVDYDPAEDNIGRSGLHMIVQARRSRQSRASALSVSGEEKSEHSEAPDHH 6562  
 QY 443 -----AYEPEEQETFEQPPPLVQQXKXSEHIDHIIQGGLSXQGLCARA 489  
 DB 6563 LSTYSIDGVAVSTAYKHATTELPOQRSSSVATQOTTSSIPSH-----PSTAGKIFRA 6617  
 QY 490 LYDYQADDDXISFDPENLITGIEVXXEGWRRGYGP---DGHGKXPANYVELI 540  
 DB 6618 MYDMADADAVSFKQGDALINVQALDEGMV--YGTVQRTGRGMLPANYVELI 6669

Search completed: April 27, 2004, 10:56:01  
 Job time : 19.097 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2004, 10:25:14 ; Search time 14.5828 Seconds  
(without alignments)  
2831.539 Million cell updates/sec

Title: US-10-028-952A-9

Perfect score: 3989  
Sequence: 1 HEIPVPTVPPAKPQXXERA.....DFRKSXXXXKLEKWTSSX 793

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	234	5.9	HS1_HUMAN	P14317 homo sapien
2	228.5	5.7	HS1_MOUSE	P49710 mus musculu
3	208	5.2	SRG8_MOUSE	Q60598 mus musculu
4	204.5	5.1	SRG8_CHICK	Q01406 gallus gall
5	202	5.1	SRG8_HUMAN	Q14247 homo sapien
6	151.5	3.8	MSRB_DICDI	P34092 dictyosteli
7	148	3.7	LSL1_MOUSE	Q61792 mus musculu
8	143.5	3.6	APB1_SCHPO	Q95768 schizosacch
9	143	3.6	APB1_YEAST	P15891 saccharomyc
10	141	3.5	ABP1_HUMAN	Q14847 homo sapien
11	134.5	3.4	ABP1_SACEX	P38479 saccharomyc
12	131	3.3	SYN1_HUMAN	P17600 homo sapien
13	129.5	3.2	MYIF_HUMAN	O00160 homo sapien
14	126	3.2	NEB1_HUMAN	P20929 homo sapien
15	124.5	3.1	ITN1_RAT	Q94074 mus musculu
16	124	3.1	ITN1_MOUSE	Q62419 mus musculu
17	123	3.1	SHJ1_MOUSE	P44522 dictyosteli
18	123	3.1	MYSC_DICDI	O88935 mus musculu
19	122.5	3.1	SYN1_MOUSE	Q63356 rattus norv
20	121.5	3.0	MYIE_RAT	Q99961 homo sapien
21	121	3.0	SHJ1_HUMAN	Q10199 schizosacch
22	120	3.0	YB2_SCHPO	O76041 homo sapien
23	120	3.0	NEB1_HUMAN	O15811 homo sapien
24	119	3.0	ITN1_HUMAN	P09991 rattus norv
25	117.5	2.9	SYN1_RAT	O61644 mus musculu
26	116.5	2.9	PAC1_MOUSE	O94075 r protein k
27	116.5	2.9	PAC1_RAT	O99111 homo sapien
28	116.5	2.9	444_1	O15735 homo sapien
29	116.5	2.9	P5PA_HUMAN	O9448 rattus norv
30	116.5	2.9	SHK1_RAT	Q63768 rattus norv
31	116	2.9	CRK_RAT	O42287 xenopus lae
32	116	2.9	PAC3_HUMAN	
33	116	2.9	ITN1_XENLA	

## ALIGNMENTS

RESULT 1	STANDARD;	PRT;	486 AA.
AC P14317;			
DT 01-JAN-1990 (Rel. 13, Created)			
DT 01-JAN-1990 (Rel. 13, Last sequence update)			
DT 15-MAR-2004 (Rel. 43, Last annotation update)			
DE Hematopoietic lineage cell specific protein (Hematopoietic cell-specific lym substrate 1) (ICKBP1).			
GN HCLSI OR HSI.			
OS Homo sapiens (Human).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX NCBI_Taxid=9606;			
RN [1]			
RP SEQUENCE FROM N.A.			
RX MEDLINE=9067934; PubMed=2587259;			
RA Kitanura D., Kaneo H., Miyagoe Y., Ariyasu T., Watanabe T.;			
RT "Isolation and characterization of a novel human gene expressed specifically in the cells of hematopoietic lineage."			
RL Nucleic Acids Res. 17:9367-9379(1989).			
RN [2]			
RP SEQUENCE FROM N.A.			
RC TTSQB=lymph;			
RX MEDLINE=23388257; PubMed=12477932;			
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F., Diatchenko L., Marstina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schetz T.E., Brownstein M.J., Udell T.B., Toshyuk S., Carninci P., Prange C., Rha S.S., Lottielano N.A., Peters G.J., Abramson R.D., Mullaly S.J., Besak S.A., McKernan P.J., McKernan K.J., Malek U.A., Gunaratne P.H., Richards S., Wooley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahey J., Hellon E., Kerteman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Buterfield Y.S.N., Krzyzanski M.I., Skalska U., Smalls D.E., Scheraga A., Schein J.E., Jones S.J.M., Marra M.A.; "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."			
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).			
RL [3]			
RP SEQUENCE OF 97-108, 193-201 AND 240-248.			
RX MEDLINE=6611348; PubMed=8713105;			
RA Egerton M., Moritz R.B., Drucker B., Kelso A., Simpson R.J., and "Identification of the 70kD heat shock cognate protein (Hsc70) and alpha-actinin-1 as novel phosphotyrosine-containing proteins in T lymphocytes."			
RT Biochem. Biophys. Res. Commun. 224:666-674(1996).			
RL [4]			
RP BINDING TO HAX-1 PROTEIN.			
RX MEDLINE=97211841; PubMed=9058608;			

34	115.5	2.9	1290	1	P1G1_RAT	P10686 rattus norv
35	115.5	2.9	1291	1	P1G1_BOVIN	P08487 bos taurus
36	114.5	2.9	486	1	PAC2_MOUSE	Q94968 mus musculu
37	114.5	2.9	488	1	PAC2_RAT	Q94917 rattus norv
38	114.5	2.9	1168	1	MYSC_ACACA	P10569 acanthamoeb
39	114	2.9	304	1	CRK_MOUSE	Q64010 mus musculu
40	114	2.9	440	1	GAGC_AYISC	P05433 avian sarc
41	114	2.9	1109	1	MYIE_HUMAN	Q12365 homo sapien
42	113.5	2.8	330	1	GRP2_HUMAN	O75791 h grb2-rela
43	113.5	2.8	452	1	YH2_YEAST	P38753 saccharomyc
44	113.5	2.8	1290	1	P1G1_HUMAN	P19174 homo sapien
45	113.5	2.8	1319	1	MN1_HUMAN	Q10571 homo sapien



RA Suzuki Y., Demoliere C., Kitamura D., Takeshita H., Deuschle U.,  
 RA Watanabe T.;  
 RT "HAX-1, a novel intracellular protein, localized on mitochondria,  
 RT directly associates with Hs1, a substrate of Src family tyrosine  
 RT kinases.";  
 RL J. Immunol. 158:2736-2744(1997).  
 RN [5]  
 RP PHOSPHORYLATION OF TYR-222.  
 RX MEDLINE=99167527; PubMed=10066823;  
 RA Brunacci A.M., Donella-Deana A., James P., Quadroni M., Contri A.,  
 RA Martin O., Pinna L.A.;  
 RT "Molecular features underlying the sequential phosphorylation of Hs1  
 RT protein and its association with c-Fgr protein-tyrosine kinase.";  
 RL J. Biol. Chem. 274:7557-7564(1999).  
 CC -1- FUNCTION: Substrate of the antigen receptor-coupled tyrosine  
 CC kinase. Plays a role in antigen receptor signaling for both  
 CC clonal expansion and deletion in lymphoid cells. Directly  
 CC associates with HAX-1, through binding to its C-terminal region.  
 CC May also be involved in the regulation of gene expression.  
 CC -1- SUBUNIT: Associates with the SH2 and SH3 domains of LCK.  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial (Probable).  
 CC -1- TISSUE SPECIFICITY: Expressed only in tissues and cells of  
 CC hematopoietic origin.  
 CC -1- DEVELOPMENTAL STAGE: Expressed in early stage of myeloid and  
 CC erythroid differentiation.  
 CC -1- PTM: Phosphorylated by LYN, rapidly after crosslinking of surface  
 CC IGM on B cells.  
 CC -1- SIMILARITY: TO CHICKEN P80/85 PROTEINS (CORTACTIN).  
 CC -1- SIMILARITY: Contains 1 SH3 domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X16663; CAA34651.1; -.  
 DR EMBL: BC016758; AAL16758.1; -.  
 DR F01: S07633; S07633.  
 DR HSP: P23355; ISEM.  
 DR Genew: HGNCG:4844; HCLSI.  
 DR MIM: 601306; -.  
 DR GO: GO:0005665; C:DNA-directed RNA polymerase II, core complex; TAS.  
 DR GO: GO:0003700; F:transcription factor activity; TAS.  
 DR GO: GO:0007242; F:intracellular signaling cascade; TAS.  
 DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; TAS.  
 DR InterPro: IPR00134; Hs1/Cortactin.  
 DR InterPro: IPR00108; Neu\_cyt\_fact\_2.  
 DR InterPro: IPR00452; SH3.  
 DR Pfam: PF02218; Hs1\_rep; 4.  
 DR Pfam: PF00018; SH3; 1.  
 DR PRINTS: PR00459; P67PHOX.  
 DR PRINTS: PR00452; SH3DOMAIN.  
 DR PRODOM: PD00006; SH3; 1.  
 DR SMART: SM00326; SH3; 1.  
 DR PROSITE: PS50002; SH3; 1.  
 KW Repeat; SH3 domain; Phosphorylation.  
 KW Repeat; SH3 domain; INVOLVED IN HAX-1 BINDING.  
 FT DOMAIN 27 66  
 FT REPEAT 81 214 3.5 X 37 AA TANDM REPEATS.  
 FT REPEAT 81 116 1.  
 FT REPEAT 117 153 2.  
 FT REPEAT 154 190 3. (INCOMPLETE).  
 FT REPEAT 191 214 4.  
 FT DOMAIN 428 486 SH3.  
 FT MOD\_RES 222 222 PHOSPHORYLATION.  
 FT MOD\_RES 241 242 KF -> FK (IN REF. 3).  
 FT CONFLICT 241 242  
 SQ SEQUENCE 486 AA; 5398 MW; 61AE637157DF5D82 CRC64;

Query Match 5.9%; Score 234; DB 1; Length 486;  
 Best Local Similarity 26.8%; Freq. No. 4.5e-12;  
 Matches 84; Conservative 25; Mismatches 97; Indels 108; Gaps 12;

QY 275 GANG--AKKXVGLXNASKESRPGKITRORGASVGLGKPKXRSPAGRP----- 325  
 DB 234 GARGLKAKPESSMAEKREKEERKAQVARROQERRA-----VTKSPSPQVIMNEEP 288  
 QY 326 --PLHAPKXGRARAPPAGTGVRCVFPFLGNGARPERPXXRPSRTARSQWTXTXS 383  
 DB 289 AVPAAPDKTSSRAMP-----VGTPSSS--SEFVTSREHPV----- 325  
 QY 384 ISSPQPKLNSPFLQQLQPTFFGRRPAAXSRADIPAEERAPSPPCU---VOAE 440  
 DB 326 -----PLLPTRQTLPEDN-----SEP-PAUPRTLEGLQVE 355  
 QY 441 EEAVERPEKQEPFYEPPPLVQOQXXGSEHIDHIQ----- 476  
 DB 356 EEPVTEAEPPPP--EEPPENDYEDVEMDMHDEDEDEPDYEVLPEPDSFSSALA 413  
 QY 477 -----GQGSXQGLCARALYDQADDXISTDPENLTIGIEVXXBGWVRGYPDG 527  
 DB 414 GSSGCPAGAGAGAVALGISAVLYDQEGSDLSFDPDVIYDIMEVDEGMWG--RCHG 472  
 QY 528 HFGKXPANYVELLE 541  
 DB 473 HFGLPANYVALLE 486  
 RESULT 2  
 Hs1\_MOUSE  
 ID Hs1\_MOUSE STANDARD; PRT; 486 AA.  
 AC P49710  
 DT 01-OCT-1996 (Rel. 34, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Hematopoietic lineage cell specific protein (Hematopoietic cell-  
 DE specific LYN substrate 1) (LCKBPI).  
 GN HCLSI OR Hs1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=95217198; PubMed=7535527;  
 RX Kitamura D., Kaneko H., Tanuchi I., Yamamura K.I., Watanabe T.;  
 RA "Molecular cloning and characterization of mouse Hs1.";  
 RL Biochem. Biophys. Res. Commun. 208:1137-1146(1995).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BA1B/c;  
 RX MEDLINE=97098949; PubMed=8943564;  
 RA Takemoto Y., Sato M., Furuta M., Hashimoto Y.;  
 RT "Distinct binding patterns of Hs1 to the Src SH2 and SH3 domains  
 RT reflect possible mechanisms of recruitment and activation of  
 RT downstream molecules.";  
 RL Int. Immunol. 8:1699-1705(1996).  
 CC -1- FUNCTION: Substrate of the antigen receptor-coupled tyrosine  
 CC kinase. Plays a role in antigen receptor signaling for both clonal  
 CC expansion and deletion in lymphoid cells. Directly associates with  
 CC HAX-1, through binding to its C-terminal region. May also be  
 CC involved in the regulation of gene expression (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Mitochondrial (By similarity).  
 CC -1- TISSUE SPECIFICITY: Expressed only in tissues and cells of  
 CC hematopoietic origin.  
 CC -1- PTM: Phosphorylated by LYN, rapidly after crosslinking of surface  
 CC IGM on B cells (By similarity).  
 CC -1- SIMILARITY: TO CHICKEN P80/85 PROTEINS (CORTACTIN).  
 CC -1- SIMILARITY: Contains 1 SH3 domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way

CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (see <a href="http://www.isb-db.ch/announce/">http://www.isb-db.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-db.ch">license@isb-db.ch</a> ).
CC	-----
DR	EMBL; D42120; BAA07701.1; -
DR	EMBL; X84797; CAA59265.1; -
DR	PIR; I49760; I49760.
DR	HASP; P29355; ISEM.
DR	MGD; MGI:104568; Hc1sl.
DR	GO; GO:0017124; F:SH3-domain binding; IDA.
DR	InterPro; IPR001334; Hel/Coactin.
DR	InterPro; IPR000108; Hel_cyt_fac2_2.
DR	InterPro; IPR01452; SH3.
DR	Pfam; PF02218; Hsl_res; 4.
DR	Pfam; PF00018; SH3_1.
DR	PRINTS; PR00499; P67PHOX.
DR	PRINTS; PR00452; SH3DOMAIN.
DR	ProDom; PD00066; SH3; 1.
DR	SMART; SMO0326; SH3; 1.
DR	PROSITE; PS0002; SH3; 1.
DR	Repeat; SH3 domain; Phosphorylation.
FT	DOMAIN 27 66 INVOLVED IN HAX-1 BINDING (BY
FT	FT 81 214 3.5 X 37 AA TANDEN REPEATS.
FT	REPEAT 81 116 1.
FT	REPEAT 117 153 2.
FT	REPEAT 154 190 3.
FT	REPEAT 191 214 4. (INCOMPLETE).
FT	DOMAIN 429 486 SH3.
FT	MOD_RRS 222 222 PHOSPHORYLATION (BY SIMILARITY).
SO	SEQUENCE 486 AA; 54212 MW; 1B5DD5D0333A0F CRC64;
Query Match	5.7%; Score 228.5; DB 1; Length 486;
Best Local Similarity	29.5%; Pred. No. 1.3e-11;
Matches	85; Conservative 24; Mismatches 122; Indels 57; Gaps 11;
QY	275 GANG-AKLXVGLXLNASKSESPRGXITRQ--RGASVGLGXPKXRLSPAGRP-PLH 328
DB	234 GARGIKAKFEELAEKREBEERKQOMARQOEKAVXKMSREVOQSMVEEPMAAQ 293
QY	329 APRKRGARAPRAGTVRCXVPVLGNGARPERPFXKRPERTASWGTWXTXISISPPQ 388
DB	294 LPKKISSEVWP-----ASHLPRESQPVRSREY-----FV 325
QY	389 PG-KLRSEFLQXLTQPTFHGREPAXXSR-PRALDPAEBAPSPCLVQAEBAVYE 446
DB	326 PSLPTRQSPLONHLEDNE----EPALPRTPEGLQVEBVEVAEAL-EPPEPDYE 379
QY	447 EPKQETVEQEPPLVQOQXGSEHIDHI-----QGQGSXQGLCARALYD 492
DB	380 PEPTEPDYEVGEIDRQDEDEGDEYDLLEBEDTPSLSYQAGPAGGAGISALALYD 439
QY	493 YQAADXKISFDPENLTIGIEVXXEGWVRGVPDGHFGKXPANVYELI 540
DB	440 YQGGSDSLSPDDIITDIEWDEGWRGQ-CRGHFGLFANVYKLL 486
RESULT 3	
SRC8_MOUSE	STANDARD; PRT; 546 AA.
AC	GQ60598;
DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	28-FEB-2003 (Rel. 41, Last annotation update)
DE	Src substrate contractin.
EN	EMSI OR CTIN.
OS	Mus musculus (Mouse).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX	NCBI_TaxId=10090;
FP	[1]
FP	SEQUENCE FROM N.A.
CC	STRAIN=BALB/c;

```

RA MEDLINE=94268839; PubMed=7516062;
RX Mclargese M.R., Mannion-Henderson J., Wu H., Parsons J.T.,
RT "the protein tyrosine kinase substrate coractin is differentially
RL expressed in murine B lymphoid tumors."
RN Oncogene 9:1989-1997(1994).
RN 121
RP SEQUENCE FROM N.P.A. AND SEQUENCE OF 125-138; 273-289 AND 534-543.
RX MEDLINE=94043284; PubMed=7693700;
RA Zhan X., Hu X., Hampton B., Burgess W.H., Friesel R., Maciag T.,
RT "Murine coractin is phosphorylated in response to fibroblast growth
RT factor-1 on tyrosine residues late in the G1 phase of the BALB/c 3T3
RT cell cycle."
RL J. Biol. Chem. 268:24427-24431(1993).
CC -1- FUNCTION: May contribute to the organization of cell structure.
CC The SH3 motif may function as a binding region to cytoskeleton.
CC Tyrosine phosphorylation in transformed cells may contribute to
CC cellular growth regulation and transformation.
CC -1- SUBUNIT: Interacts with SHANK2 and SHANK3 via its SH3 domain (By
CC similarity).
CC -1- TISSUE SPECIFICITY: Detected in most murine tissues, but not
CC detected in B lymphocytes or plasma cells.
CC -1- SIMILARITY: TO HEMATOPOIETIC LINEAGE CELL SPECIFIC PROTEIN.
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcement/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; U03184; AAA19689.1; -.
DR PIR; I48899; I48899.
DR HSSP; P19174; 2HSP.
DR MED; MG1:99685; Cctn.
DR InterPro; IPR001334; Hs1/coractin.
DR InterPro; IPR000108; Neu_cyt_fac2_2.
DR InterPro; IPR001452; SH3.
DR Pfam; PF02218; Hs1_rep; 7.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PRO0499; P67PHOX.
DR PRINTS; PRO0452; SH3DOMAIN.
DR PRODOM; ED000066; SH3; 1.
DR SMART; SM00326; SH3; 1.
DR DR PROSITE; PS50002; SH3; 1.
DR Phosphorylation; Repeat; SH3 domain; Cytoskeleton.
DR DOMAIN 83 329 6.5 X 37 AA TANDEN REPEATS.
FT REPEAT 83 119 1.
FT REPEAT 120 156 2.
FT REPEAT 157 193 3.
FT REPEAT 194 230 4.
FT REPEAT 231 267 5.
FT REPEAT 268 304 6.
FT REPEAT 305 329 7 (INCOMPLETE).
FT DOMAIN 488 546 SH3.
FT CONFLICT 9 9 A -> R (IN REF. 2).
SO SEQUENCE 546 AA; 61260 MW; 8F93A026AD1D6D4F CMC64;

Query Match 5.24; Score 208; DB 1; Length 546;
Best Local Similarity 35.6%; Pred. No. 7.4e-10;
Matches 57; Conservative 18; Mismatches 57; Indels 28; Gaps 6

QY 386 SPQGRKRSRPLQQLQTPETHFRREPAAKXSRPADPAEPAPASXPCVQAEAEAY 445
DB 407 SPQPIEDRP--SSPIYEDAAPFALESYRSEF-----EPEYS-----IEAGIP 450
QY 446 EEPXQETFEQRLPVQOQQXGSEHIDHIIQGG-----ISKGLCARALYDQADDX 499
DB 451 EAGSQGGGLTSEVYE-----TEAFGHYQADPDYDYGESDIDGITALYDQADBD 505
QY 500 EISFDPENLITGIEVXKGMWRGQVGGDGHFQMXANAYEL 539

```

Db 506 EISFDPDITINEMIDGMRGV-CKGRYGLFPANVEL 544

RESULT 4

SC8B\_CHICK STANDARD; PRT; 563 AA.

AC Q01406; 01-APR-1993 (Rel. 25, Created)

DT 01-APR-1993 (Rel. 25, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Src substrate protein p85 (p80) (Corrctin).

GN Ems1 OR p85.25.

OS Gallus gallus (Chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.

OC NCBI\_TaxID=9031;

OX [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=92017786; PubMed=1922035; Wu H., Reynolds A.B., Kanner S.B., Vines R.R., Parsons J.T.; "Identification and characterization of a novel cytoskeleton-associated p85src substrate." Mol. Cell. Biol. 11:5113-5124(1991).

RT Mol. Cell. Biol. 11:5113-5124(1991).

CC -1- FUNCTION: May contribute to the organization of cell structure. The SH3 motif may function as a binding region to cytoskeleton. Tyrosine phosphorylation in transformed cells may contribute to cellular growth regulation and transformation.

CC -1- SUBCELLULAR LOCATION: In normal cells, probably in association with the plasma membrane and possibly the endoplasmic reticulum. p80/85 colocalizes with F-actin in peripheral extensions of normal cells and reconstitutes (podosomes) of src-transformed cells. p80/85 probably associates with components of the cytoskeleton.

CC -1- ALTERNATIVE PRODUCTS:

CC Event=Alternative initiation;

CC Comment=2 isoforms, p85 (shown here) and p80, may be produced by alternative initiation;

CC -1- PTM: In normal cells, appears to be phosphorylated on serine and threonine; in cells expressing activated forms of pp60-src, they become heavily phosphorylated on tyrosine in vitro.

CC -1- SIMILARITY: TO HEMATOPOIETIC LINEAGE CELL SPECIFIC PROTEIN.

CC -1- SIMILARITY: Contains 1 SH3 domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M73705; AAA49031.1; -

DR PIR; A41530; A41530.

DR HSSP; P19174; 2HSP.

DR InterPro; IPR000134; Hsl/Corrctin.

DR InterPro; IPR000108; Neu\_Cyc\_fact\_2.

DR InterPro; IPR001452; SH3.

DR Pfam; PF02218; Hsl\_rep; 7.

DR Pfam; PF00018; SH3; 1.

DR PRINTS; PR00499; P67PHOX.

DR PRINTS; PR00452; SH3DOMAIN.

DR PRODOM; PD000066; SH3; 1.

DR SMART; SM00326; SH3; 1.

DR PROSITE; PS50002; SH3; 1.

DR Phosphorylation; Repeat; SH3 domain; Cytoskeleton; KX Alternative initiation.

MM Alternative initiation.

FT CHAIN 1 563 SRC SUBSTRATE PROTEIN P85, ISOFORM P85.

FT INT MET 11 11 FOR ISOFORM P80.

FT DOMAIN 92 338 6.5 X 37 AA TANDEN REPEATS.

FT REPEAT 92 128 1.

FT REPEAT 129 165 2.

FT REPEAT 166 202 3.

FT REPEAT 203 239 4.

FT REPEAT 240 276 5.

FT REPEAT 277 313 6.

FT REPEAT 314 338 7 (INCOMPLETE).

FT REPEAT 338 353 8.

FT DOMAIN 505 563 SH3.

SQ SEQUENCE 563 AA; 63329 MW; 22A824A08B8D54C CRC64;

Query Match 5.1%; Score 204.5; DB 1; Length 563;

Best Local Similarity 36.1%; Pred. No. 1.5e-09;

Matches 52; Conservative 18; Mismatches 45; Indels 29; Gaps 5;

Db 424 PAEPAPSPX--PCLVQAEAEVYEPX-----EQETFE-----QPLVQOQXKS 468

419 PAEPKPPSPVQDANVSAYASAYKKSSTTYSAEHPESGCTGSDYQEAVSOREAEYE 478

QY 469 EVIDHHIQOQGLSXQ-----GLCARLYDQADDXEISPDENLITGIEYX 515

Db 479 PETTYEVAGAGDHYQAEENTYDEYENELGITALYDQAAGDEISFDPDITINEMI 538

QY 516 XEGMWRGYGPDGHFGKXPANVEL 539

Db 539 DDCMWRGV-CKGRYGLFPANVEL 561

# RESULT 5

SC8B\_HUMAN STANDARD; PRT; 550 AA.

ID SC8B\_HUMAN 014247;

DT 01-NOV-1997 (Rel. 35, Created)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Src substrate cortactin (Amplixin) (Oncogene Ems1).

GN Ems1 OR CTN.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OC NCBI\_TaxID=9606;

OX [1]

RP SEQUENCE FROM N.A.

RX TISSUE=Breast;

RX MEDLINE=92195667; PubMed=1532244;

RA Schuring E.M.D., Verhoeven B., Wool W.J., Michalides R.J.A.; "Identification and cloning of two overexpressed genes, U21B3/PRAD1 and Ems1, within the amplified chromosome 11q13 region in human carcinoma." Oncogene 7:355-361(1992).

RT (2)

RP SEQUENCE FROM N.A.

RX TISSUE=Breast;

RX MEDLINE=9233650; PubMed=8474448;

RA Schuring E.M.D., Verhoeven B., Litvinov S., Michalides R.J.A.; "The product of the Ems1 gene, amplified and overexpressed in human carcinoma, is homologous to a v-src substrate and is located in cell-substratum contact sites." Cell 61:13-28(1993).

RT Mol. Cell. Biol. 13:2891-2898(1993).

CC -1- FUNCTION: May contribute to the organization of cell structure. Tyrosine phosphorylation in transformed cells may contribute to cellular growth regulation and transformation.

CC -1- SUBUNIT: Interacts with SHANK2 and SHANK3 via its SH2 domain (By similarity).

CC -1- SIMILARITY: TO HEMATOPOIETIC LINEAGE CELL SPECIFIC PROTEIN.

CC -1- SIMILARITY: Contains 1 SH3 domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M73705; AAA49031.1; -

DR PIR; A41530; A41530.

DR HSSP; P19174; 2HSP.

DR InterPro; IPR000134; Hsl/Corrctin.

DR InterPro; IPR000108; Neu\_Cyc\_fact\_2.

DR InterPro; IPR001452; SH3.

DR Pfam; PF02218; Hsl\_rep; 7.

DR Pfam; PF00018; SH3; 1.

DR PRINTS; PR00499; P67PHOX.

DR PRINTS; PR00452; SH3DOMAIN.

DR PRODOM; PD000066; SH3; 1.

DR SMART; SM00326; SH3; 1.

DR PROSITE; PS50002; SH3; 1.

DR Phosphorylation; Repeat; SH3 domain; Cytoskeleton; KX Alternative initiation.

MM Alternative initiation.

FT CHAIN 1 563 SRC SUBSTRATE PROTEIN P85, ISOFORM P85.

FT INT MET 11 11 FOR ISOFORM P80.

FT DOMAIN 92 338 6.5 X 37 AA TANDEN REPEATS.

FT REPEAT 92 128 1.

FT REPEAT 129 165 2.

```

DR EMBL; M98343; AAA58455.1; --
DR PIR; A48063; A48063.
DR HSSP; P19174; 2HSP.
DR Genew; HGNC:3338; EMS1.
DR MIM; 164765; --
DR GO; GO:0005856; C:cytoskeleton; TAS.
DR GO; GO:0005625; C:soluble fraction; TAS.
DR InterPro; IPR003134; Hsl/Contractin.
DR InterPro; IPR00108; Neu_cyt_fact_2.
DR InterPro; IPR001452; SH3.
DR Pfam; PF02218; Hsl_rep; 7.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00459; P67PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PSS0002; SH3; 1.
DR Phosphorylation; Repeat; SH3 domain; Cytoskeleton.
DR DOMAIN 83 329 6.5 X 37 AA TANDEN REPEATS.
FT REPEAT 83 119 1.
FT REPEAT 120 156 2.
FT REPEAT 157 193 3.
FT REPEAT 194 230 4.
FT REPEAT 231 267 5.
FT REPEAT 268 304 6.
FT REPEAT 305 329 7 (INCOMPLETE).
FT DOMAIN 492 550 SH3.
SQ SEQUENCE 550 AA; 61636 MW; 2799227C2B4383A7 CRC64;

Query Match 5.1%; Score 202; DB 1; Length 550;
Best Local Similarity 36.0%; Pred. No. 2.3e-09;
Matches 58; Conservative 17; Mismatches 60; Indels 26; Gaps 6;

QY 386 SPQPKLSPFLQXLTQPTETFGEPAXXSRPADIPAEBAESPXPCLVQAEBAVY 445
DB 407 APQPEERLP-----SSPV-----YEDASF--KAELSYGVPVSGTPEPVYSMEADY 453
QY 446 EEPXQ-----EFFYEPPLVQOQXGSEIHDIHIOGGLSXLGLCARLYYQAD 498
DB 454 REASSQGLAYATEVAVE---SAEAGHYPAEDSTDEYENDLGYTVALYDQAGD 508
QY 499 XEISFPENLITGIEVXXEGMWRGYGPDGHFGMKXPANYVEL 539
DB 509 DEISFPDIIITIMIDGMWRGV-CGGRYGLFPANVEL 548

RESULT 6
MYSB_DICDI STANDARD; PRT; 1111 AA.
AC P34052;
DT 01-FEB-1994 (Rel. 28. Created)
DT 01-FEB-1994 (Rel. 28. Last sequence update)
DT 15-MAR-2004 (Rel. 43. Last annotation update)
DE MYOSIN IB heavy chain.
GN MYOB OR MYOA OR DMIB.
OS Dictyostelium discoideum (Slime mold).
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX NCBI_Taxid=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AX3;
RX MEDLINE=89345628; PubMed=2762320;
RA Jung G., Saxe C.L. II, Kimmel A.R., Hammer J.A. III;
RT "Dictyostelium discoideum contains a gene encoding a myosin I heavy
chain."
RL Proc. Natl. Acad. Sci. U.S.A. 86:6186-6190(1989).
RN [2]
RP SEQUENCE OF 481-490; 656-666 AND 783-798.
RC STRAIN=AX3;
RX MEDLINE=93315475; PubMed=8325874;
RA Jung G., Fukui Y., Martin B., Hammer J.A. III;
RT "Sequence, expression pattern, intracellular localization, and
targeted disruption of the Dictyostelium myosin I heavy chain

```

```

RT isoform.";
RL J. Biol. Chem. 268:14981-14990(1993).
CC -1- FUNCTION: Myosin is a protein that binds to actin and has ATPase
activity that is activated by actin. Myosin IB may have a role in
chemotaxis and aggregation; it could serve to stabilize and even
retract cortical structures, such as pseudopods and lamellipods.
CC -1- SUBUNIT: Myosin I heavy chain is single-headed. Dimer of a heavy
and a light chain. Inability to self-assemble into filaments.
CC -1- SUBCELLULAR LOCATION: Highest concentration just beneath the
plasma membrane in the anterior pseudopod at the leading edge of
the cell.
CC -1- DOMAIN: TH.1 binds directly to anionic phospholipid membranes;
myosins I could therefore move actin relative to membranes and
vice versa. TH.2 and SH3 bind tightly to F-actin; this together
with the nucleotide-sensitive site in the head, allows single
molecules of myosin I to cross-link actin filaments.
CC -1- SIMILARITY: Contains 1 myosin-like globular head domain.
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
DR EMBL; M26037; AAA33229.1; --
DR PIR; A33284; A33284.
DR HSSP; P08799; 1MND.
DR DictyBase; DDB0002013; MYOB.
DR InterPro; IPR000048; IQ region.
DR InterPro; IPR001609; myosin_head.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00612; IQ; 1.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR PRINTS; PR00452; SH3DOMAIN.
DR ProDom; PD000355; myosin_head; 1.
DR ProDom; PD000066; SH3; 1.
DR SMART; SM00242; MISC; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PSS0002; SH3; 1.
KM Myosin; Actin-binding; ATP-binding; SH3 domain; Multigene family;
KM Chemotaxis; Phosphorylation.
FT DOMAIN 1 694 MYOSIN HEAD-LIKE.
FT DOMAIN 695 921 TAIL HOMOLOG REGION 1 (TH.1).
FT DOMAIN 922 1052 GLY/PRO/ALA-RICH (TH.2).
FT DOMAIN 1053 1111 SH3.
FT NP_BIND 102 109 ATP (POTENTIAL).
FT DOMAIN 547 627 ACTIN-BINDING.
FT DOMAIN 951 1015 ASN-RICH.
FT MOD_RES 332 332 PHOSPHORYLATION (BY SIMILARITY).
SQ SEQUENCE 1111 AA; 124313 MW; CD6386F08DC642F CRC64;

Query Match 3.8%; Score 151.5; DB 1; Length 1111;
Best Local Similarity 21.7%; Pred. No. 7.8e-05;
Matches 57; Conservative 21; Mismatches 86; Indels 99; Gaps 5;

QY 291 SKSESPPGKXIRRGASVGLGXKXRLSPAPRPPLHAPRXGRAPPAAGTGYRCV 350
DB 931 SQASSRFA---PQSGRGRRGPG-----AQPPRPQDQDQKFMPPRQGGGARRP 981
QY 351 PFLVNGARPERPRXPRSTARSWTGXYSISPPQPKLSPFLQXLTQPTETFG 410
DB 982 OPQGGGARP-----MGAPQGGARQDQAGRGQLPPTQGGGA 1018
QY 411 E-----PAXXSRPADIPAEBAESPXPCLVQAEBAVYEEKQETFY 455
DB 1019 PGGGAPMGSGAPGGGAGGAGPPLPTVAKPAPQRP----- 1056
QY 456 EPPPLVQOQXGSEIHDIHIOGGLSXLGLCARLYYQADDEXISFPENLITGIEVX 515

```

RA EMBL; U58882; AAC52639.1; -.  
DR HSSP; P60171; 1ZFO.  
DR MGD; MG1.109566; LaeP1.  
DR InterPro; IPR001781; LIM.  
DR InterPro; IPR003900; Nebulin.  
DR InterPro; IPR001452; SH3.  
DR Pfam; PF00412; LIM; 1.  
DR Pfam; PF00880; Nebulin; 2.  
DR Pfam; PF00018; SH3; 1.  
DR PRINTS; PR00452; SH3DOMAIN.  
DR ProDom; PD000094; LIM; 1.  
DR ProDom; PD000066; SH3; 1.  
DR SMART; SM00132; LIM; 1.  
DR SMART; SM00227; NEBU; 2.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PS00478; LIM\_DOMAIN\_1; 1.  
DR PROSITE; PS00023; LIM\_DOMAIN\_2; 1.  
DR PROSITE; PS00023; SH3; 1.  
DR LIM domain; Metal-binding; Zinc; SH3 domain.  
FT DOMAIN 5 56 LIM.  
FT DOMAIN 204 263 SH3.  
FT DOMAIN 174 180 POLY-GLN.  
FT DOMAIN 203 206 POLY-GLY.  
SQ SEQUENCE 263 AA; 29994 MM; A6CA2FC2E451433E CRC64;  
Query Match 3.7%; Score 148; DB 1; Length 263;  
Best Local Similarity 33.1%; Pred. No. 3.3e-05;  
Matches 41; Conservative 16; Mismatches 41; Indels 26; Gaps 5  
QY 438 QAEAEAVYEPPEDE-----TFYEQPPLVQOQXXGSHIHIIQGLSXQ--- 483  
DB 145 EADSSSYRPFQEQQPQPHHIFPSAPYQDP---QQQMTSSYGKYPAPVSIQRA 201  
QY 484 ----GHCARALDYQPADDXEISFDPENLITGIEVXXGWRGYP---DGFQKXPANY 536  
DB 202 PGCGGRYRAYVDYSADDEDEVSFDGDTIVVQIDGWM--YGVTERYGTGMLPANY 259  
QY 537 VEI 540  
DB 260 VEI 263  
RESULT 8  
APPL\_SCHPO STANDARD; PRT; 857 AA.  
ID APPL\_SCHPO  
AC Q9P7E8; F78815; 42, Created  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DR APPL protein.  
GN APPL OR SPAPJ760.02C.  
OS Schizosaccharomyces pombe (fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=21848401; PubMed=11859360;  
RA Wood V., Gwilliam R., Rajadream M.A., Lyne M., Lyne R., Stewart A.,  
Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
Holtroyd S., Hornsby T., Howarth S., Huckle E.U., Hunt S., Jagsels K.,  
James K., Jones L., Jones M., Leather S., McDonald S., McLean U.,  
Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
Skellern J., Simmonds M., Squares R., Stevens K.,  
Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
Woodward J., Volckaert G., Aert R., Robben J., Grynolpriez B.,

RA Weljens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehmach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Weiler H., Mambrot R., Punelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaive V., Mottier S.,  
 RA Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerutti L., Lowe T., McCombie M.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;  
 RA "The genome sequence of *Schizosaccharomyces pombe*.";  
 RT Nature 415:871-880(2002).  
 RN [2]  
 RP SEQUENCE OF 560-857 FROM N.A.  
 RC STRAIN=PR745;  
 RX MEDLINE=98162722; PubMed=9501991;  
 RA Yoshida S., Kato K., Nakai K., Okayama H., Nojima H.;  
 RT "Identification of open reading frames in *Schizosaccharomyces pombe*  
 RT CDNA.";  
 RL DNA Ref. 4:363-369(1997).  
 CC -1- SIMILARITY: Contains 2 SH3 domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AL162631; CAB83085.1; -;  
 DR EMBL: D89164; BAAL3826.1; ALT\_SEQ.  
 DR HSRP: P19174; SH3P.  
 DR Genedb\_Spomer; SPAB0760.02c; -;  
 DR InterPro; IPR002108; Actbind\_cofl1.  
 DR InterPro; IPR001452; SH3.  
 DR Pfam; PF00241; cofilin\_ADF; 1.  
 DR Pfam; PF00018; SH3; 2.  
 DR PRINTS; PR00499; P67PHOX.  
 DR PRINTS; PR00452; SH3DOMAIN.  
 DR PRODOM; PD000066; SH3; 2.  
 DR SMART; SM00102; ADF; 1.  
 DR SMART; SM00326; SH3; 2.  
 DR PROSITE; PS50002; SH3; 2.  
 KW SH3 domain; Repeat.  
 FT DOMAIN 725 785 SH3 1.  
 FT 860 857 SH3 2.  
 SQ SEQUENCE 857 AA; 91137 MW; 247EFA78BAFE7097 CRC64;

Query Match 3.6%; Score 143.5; DB 1; Length 857;  
 Best Local Similarity 27.3%; Pred. No. 0.00027;  
 Matches 51; Conservative 25; Mismatches 60; Indels 51; Gaps 10;

QY 387 PQGKLRPFLQXLTQETFRGRPRAXXSRPADLPA-EPAPSAKPCVQAEBAVY 445  
 DB PEAQGLNEPVPPLPHDET--QEPQVG-----GDVATHTOPTKTPALV-----IY 734  
 QY 446 EEPXQETFEYEQPLVQ-QQXXGSEHD-----HHIQG-QGL----- 480  
 DB 735 DYSPEENHEIF---LVNEQIQLFVDDGWLGNSSGGGLPPSNVETITGNEIANN 791  
 QY 481 -----SXQGLCAALADYQADDXETSPDENITIGIYVXXRGWRGVPDGHFGKXP 533  
 DB 792 PPAEPQAGPGKSVKAIYDQKQEDNELSFEDIELIANVDCVDPWMEG-ECHGRGLFP 850  
 QY 534 ANYVELI 540  
 DB 851 SNYVERI 857

ABP1 YEAST  
 ID ABP1 YEAST STANDARD; PRT; 592 AA.  
 AC P15851;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DS Actin binding protein.  
 GN ABP1 OR YCR088W OR YCR88W.  
 OS *Saccharomyces cerevisiae* (Baker's yeast).  
 CC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 CC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.  
 CX NCBI\_TaxID=4932;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90136906; PubMed=2405279;  
 RA Drubin D.G., Mulholland J., Zhu Z., Botstein D.;  
 RT "Homology of a yeast actin-binding protein to signal transduction  
 RT proteins and myosin-I.";  
 RL Nature 343:288-290(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Frontali L., Grisanti P.;  
 RL Submitted (MAR-1992) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: MAY BE INVOLVED IN THE SPATIAL ORGANIZATION OF CELL  
 CC SURFACE GROWTH. AN OVERPRODUCTION OF ABP1 CAUSES THE ASSEMBLY OF  
 CC THE CORTICAL ACTIN SKELETON AT INAPPROPRIATE SITES ON THE CELL  
 CC SURFACE, RESULTING IN DELOCALIZED SURFACE GROWTH.  
 CC -1- SUBCELLULAR LOCATION: CORTICAL CYTOSKELETON.  
 CC -1- SIMILARITY: TO SIGNAL TRANSDUCTION PROTEINS AND MYOSIN I.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X51780; CAA36075.1; -;  
 DR EMBL: X59720; CAA42253.1; -;  
 DR PIR; S19503; LBY1.  
 DR PDB; 1HQ2; 21-DEC-01.  
 DR PDB; 1Y08; 01-MAR-02.  
 DR Germonline; 138985; -;  
 DR SGD; S0000684; ABP1.  
 DR GO; GO:0005938; C:cell cortex; IDA.  
 DR GO; GO:0005737; C:cytoplasm; IDA.  
 DR GO; GO:0005515; F:protein binding; IDA.  
 DR GO; GO:0000147; P:actin cortical patch assembly; IMP.  
 DR InterPro; IPR002108; Actbind\_cofl1.  
 DR InterPro; IPR000108; Neu\_cyt\_fact\_2.  
 DR InterPro; IPR001452; SH3.  
 DR Pfam; PF00241; cofilin\_ADF; 1.  
 DR Pfam; PF00018; SH3; 1.  
 DR PRINTS; PR00499; P67PHOX.  
 DR PRINTS; PR00452; SH3DOMAIN.  
 DR PRODOM; PD000066; SH3; 1.  
 DR SMART; SM00102; ADF; 1.  
 DR SMART; SM00326; SH3; 1.  
 DR PROSITE; PS50002; SH3; 1.  
 KW Cytoskeleton; Actin-binding; SH3 domain; Repeat. 3D-structure.  
 FT DOMAIN 86 97  
 FT 532 592 SH3.  
 FT DOMAIN 200 575 3 x 10 AA APPROXIMATE REPEATS.  
 FT REPEAT 200 209 1.  
 FT REPEAT 436 445 2.  
 FT REPEAT 456 465 3.  
 FT REPEAT 566 575 3.  
 FT REPEAT 58 58 1 -> S (IN REF. 1).  
 FT CONFLICT 312 312 K -> I (IN REF. 1).  
 FT SEQUENCE 592 AA; 65576 MW; 39523510704D94AA CRC64;

Query Match 3.6%; Score 143; DB 1; Length 592;  
 Best Local Similarity 27.2%; Pred. No. 0.0002;  
 Matches 56; Conservative 21; Mismatches 79; Indels 50; Gaps 8;

QY SSSPPGKRSRPLQXQLOPETHFR-----EP-----412  
 DB TTSKPSKQEPKAEAPKTDYKKNPLPGMHLEADNEEPENDDDWDEDE 445  
 QY 383 SSSPPGKRSRPLQXQLOPETHFR-----EP-----412  
 DB TTSKPSKQEPKAEAPKTDYKKNPLPGMHLEADNEEPENDDDWDEDE 445  
 QY 413 AAXXSRPR-----ADLPAEPP-----APSYR-----PCLVAEEBAYEEPEXQET 454  
 DB 446 AAOPLPSRNVASGAPVQKEPEDEIAPSLPSRNSIPAPQAEAPQAP-EEIEEEAE 504  
 QY 455 YEQPLVQGXGSHIDHIIQGGLSXQGLCAPALYDCAADXEISFPDENLITGIEV 514  
 DB 505 EAPPLPSRSSASAPPPPPRRATPEKKPKENPMWAIHEDYDAEDNEELTFVNDKINIEF 564  
 QY 515 XXEGWVRG-YGPDGHFGKXPANYVEL 539  
 DB 565 VDDDWMLGELKXGSKGLFPSNYVEL 590

RESULT 10  
 LAST\_HUMAN STANDARD; PRT; 261 AA.

AC Q14847; Q96ED2; 36, Created)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 15-JUL-1998 (Rel. 36, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE LIM and SH3 domain protein 1 (LASP-1) (MLN 50).  
 GN LASP1 OR MLN50.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID:9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=breast carcinoma;  
 RX MEDLINE=96039245; PubMed=7490069;  
 RA Lidereau R., Bassot P., Rio M.-C.;  
 RA "Identification of four novel human genes amplified and overexpressed  
 in breast carcinoma and localized to the q11-q21.3 region of  
 chromosome 17.";  
 RT Genomics 28:367-376 (1995).  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Liver;  
 RX MEDLINE=2388257; PubMed=12477932;  
 RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Suetow K.H., Scheffer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Stachenko L., Marusina K., Farmer A.A., Rubin G.W., Hong L.,  
 RA Diaperon M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.B.,  
 RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Mullany S.J.,  
 RA Raha S.S., Lomeli N.A., Peters G.J., Abramson R.D., Mullany S.J.,  
 RA Bosak S.A., McMan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richardson S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Muzny K.C., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey U., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Blakeley A.C., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywicki M.I., Skalek U., Small D.B.,  
 RA Scherch A., Schen U.E., Jones S.J.W., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [3]  
 RP DOMAINS.  
 RX MEDLINE=96033982; PubMed=7589475;  
 RA Tomasetto C., Moog-Lutz C., Regnier C.H., Schreiber V., Bassot P.,  
 RA Rio M.-C.;

RT "Lasp-1 (MLN 50) defines a new LIM protein subfamily characterized by  
 the association of LIM and SH3 domains.";  
 RL FEBS Lett. 373:245-249 (1995).  
 CC -1- SIMILARITY: Contains 1 LIM zinc-binding domain.  
 CC -1- SIMILARITY: Contains 1 SH3 domain.  
 CC -1- DATABASE: NAME=Atlas Genet. Cytoenet. Oncol. Haematol.;  
 WWW="http://www.infobiogen.fr/services/chronocance/genes/lasp1ID203.html".  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb.ch/announce/>  
 or send an email to [license@isb.ch](mailto:license@isb.ch)).  
 CC

DR EMBL; X82456; CA457833.1; -  
 DR EMBL; BC012460; AAH12460.1; -  
 DR HSP; P80171; IZFO.  
 DR SWISS-2DPAGE; Q14847; HUMAN.  
 DR Genew; HGNC:6513; LASP1.  
 DR MIM; 602920; -  
 DR GO; GO:0005070; F:SH3/SH2 adaptor protein activity; TAS.  
 DR InterPro; IPR001781; LIM.  
 DR InterPro; IPR000900; Nebulin.  
 DR InterPro; IPR001452; SH3.  
 DR Pfam; PF00412; LIM; 1.  
 DR Pfam; PF00880; Nebulin; 2.  
 DR Pfam; PF00018; SH3; 1.  
 DR PRINTS; PR00453; SH3DOMAIN.  
 DR Prodom; PD000064; LIM; 1.  
 DR Prodom; PD000066; SH3; 1.  
 DR Prodom; PD000132; LIM; 1.  
 DR SMART; SM00326; SH3; 1.  
 DR SMART; SM00227; NEBU; 2.  
 DR PROSITE; PS00478; LIM DOMAIN\_1; 1.  
 DR PROSITE; PS50023; LIM DOMAIN\_2; 1.  
 DR PROSITE; PS50002; SH3; 1.  
 DR LIM domain; Metal-binding; Zinc; SH3 domain.  
 FT DOMAIN 5 36  
 FT DOMAIN 202 261  
 FT DOMAIN 201 204  
 FT CONFLICT 79 79  
 FT CONFLICT 210 210  
 FT CONFLICT 220 220  
 SQ SEQUENCE 261 AA; 29717 MW; 3B89B988605B3639 CRC64;

Query Match 3.5%; Score 141; DB 1; Length 261;  
 Best Local Similarity 29.9%; Pred. No. 0.00013;  
 Matches 44; Conservative 17; Mismatches 48; Indels 38; Gaps 7;

QY 409 GREPAXXSRP-----RADLPAEPP-----APSXPCLVQAEBAV-----YEEPEQETFE 456  
 DB 138 GMEBERRSDGSGSYRPLAQCPHIFTSAPVQCPQCPVQAGVGYKEPA-----190  
 QY 457 QPPLVQGXGSHIDHIIQGGLSXQGLCAPALYDCAADXEISFPDENLITGIEV 516  
 DB 191 APVSIQASABG-----GKRYRAYDYASABDEVSFGDGTIVNQOID 236  
 QY 517 EGMWVRGYP---DGHGKXPANYVEL 540  
 DB 237 GGMW-YGVTERIGDTGMLPANYVEL 261

RESULT 11  
 ABPL\_SACEX STANDARD; PRT; 617 AA.  
 DT 01-OCT-1994 (Rel. 30, Created)  
 DT 01-OCT-1994 (Rel. 30, Last sequence update)  
 DT 01-FEB-1995 (Rel. 31, Last annotation update)  
 DE Actin binding protein.  
 GN ABPL.



```

OS Saccharomyces exiguus (yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCB1_TaxID=34358;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CBS 379;
RX MEDLINE=9415401; PubMed=8110838;
RA Lange U., Steiner S., Grolig F., Wagner G., Philippsen P.;
RT "Cloning and sequencing of a gene coding for an actin binding protein
RT of Saccharomyces exiguus."
RL Biochim. Biophys. Acta 1217:214-218(1994).
CC -1- FUNCTION: MAY BE INVOLVED IN THE SPATIAL ORGANIZATION OF CELL
CC SURFACE GROWTH. AN OVERPRODUCTION OF ABP1 CAUSES THE ASSEMBLY OF
CC THE CORTICAL ACTIN SKELETON AT INAPPROPRIATE SITES ON THE CELL
CC SURFACE, RESULTING IN DELOCALIZED SURFACE GROWTH (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: CORTICAL CYTOSKELETON.
CC -1- SIMILARITY: TO SIGNAL TRANSDUCTION PROTEINS AND MYOSIN I.
CC -1- SIMILARITY: Contains 1 SH3 domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X73977; CAA52156.1; -
DR F1R; S42719; S42719.
DR HSSP; P19174; 2HSP.
DR InterPro; IPR002108; Actbind_cofln.
DR InterPro; IPR000108; Neu_cyt_fact_2.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00241; Cofilin_ADF; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00459; P67PHOX.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRODOM; PD000066; SH3; 1.
DR SMART; SM00102; ADF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS00002; SH3; 1.
DR Cytoskeleton; Actin-binding; SH3 domain; Repeat.
KW DOMAIN
FT 86 97
FT DOMAIN
FT 557 617
FT DOMAIN
FT 486 501
FT DOMAIN
FT 523 529
FT DOMAIN
FT 202 600
FT REPEAT
FT 202 211
FT REPEAT
FT 444 453
FT REPEAT
FT 581 600
FT DOMAIN
FT 495 538
FT REPEAT
FT 495 510
FT REPEAT
FT 523 538
FT REPEAT
FT 617 640
SQ SEQUENCE 617 AA; 68420 MW; 540A785191B65F85 CRC64;

Query Match 3.4%; Score 134.5; DB 1; Length 617;
Best Local Similarity 25.0%; Pred. No. 0.0011;
Matches 52; Conservative 21; Mismatches 82; Indels 53; Gaps 8;

```

```

QY 513 EVXXEGWNRG-YGPDGHFGKXPANYVEL 539
DB 588 EFVDDWMTLGELETTGQKGLFPSNYVL 615

RESULT 12
ID SYN1_HUMAN STANDARD; PRT; 705 AA.
AC P17600; 075825;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Synapsin I (Brain protein 4.1).
GN SYN1
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCB1_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=90243651; PubMed=2110562;
RA Suedhof T.C.;
RT "The structure of the human synapsin I gene and protein."
RL J. Biol. Chem. 265:7849-7852(1990).
RN [2]
RP SEQUENCE OF 1-125 FROM N.A.
RX MEDLINE=90368667; PubMed=2118519;
RA Sauerwald A., Hoesche C., Oeschwald R., Kilmann M.W.;
RT "The 5'-flanking region of the synapsin I gene. A G-C-rich, TATA- and
RT CAT-less, phylogenetically conserved sequence with cell
RT type-specific promoter function."
RL J. Biol. Chem. 265:14932-14937(1990).
RN [3]
RP SEQUENCE OF 1-258 FROM N.A.
RA Graefham D.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Neuronal phosphoprotein that coats synaptic vesicles,
CC binds to the cytoskeleton, and is believed to function in the
CC regulation of neurotransmitter release.
CC -1- SUBCELLULAR LOCATION: Synapse.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=Synapsin IA;
CC IsoId=P17600-1; Sequence=Displayed;
CC Name=Synapsin IB;
CC IsoId=P17600-2; Sequence=VSP_006316, VSP_006317;
CC -1- PTM: SUBSTRATE OF AT LEAST FOUR DIFFERENT PROTEIN KINASES. IT IS
CC PROBABLE THAT PHOSPHORYLATION PLAYS A ROLE IN THE REGULATION OF
CC SYNAPSIN I IN THE NERVE TERMINAL.
CC -1- SIMILARITY: Belongs to the synapsin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M58378; AAC41930.1; -
DR EMBL; M58321; AAC41930.1; JOINED.
DR EMBL; M58341; AAC41930.1; JOINED.
DR EMBL; M58351; AAC41930.1; JOINED.
DR EMBL; M58353; AAC41930.1; JOINED.
DR EMBL; M58359; AAC41930.1; JOINED.
DR EMBL; M58371; AAC41930.1; JOINED.
DR EMBL; M58372; AAC41930.1; JOINED.
DR EMBL; M58373; AAC41930.1; JOINED.
DR EMBL; M58374; AAC41930.1; JOINED.
DR EMBL; M58375; AAC41930.1; JOINED.
DR EMBL; M58376; AAC41930.1; JOINED.
DR EMBL; M58377; AAC41930.1; JOINED.

```

DR	EMBL/ M56378; AAC41931.1; ALT SEQ.
DR	EMBL/ M56321; AAC41931.1; JOINED
DR	EMBL/ M56341; AAC41931.1; JOINED
DR	EMBL/ M56351; AAC41931.1; JOINED
DR	EMBL/ M56353; AAC41931.1; JOINED
DR	EMBL/ M56359; AAC41931.1; JOINED
DR	EMBL/ M56372; AAC41931.1; JOINED
DR	EMBL/ M56372; AAC41931.1; JOINED
DR	EMBL/ M56374; AAC41931.1; JOINED
DR	EMBL/ M56375; AAC41931.1; JOINED
DR	EMBL/ M56376; AAC41931.1; JOINED
DR	EMBL/ M56377; AAC41931.1; JOINED
DR	EMBL/ M56301; AAA60608.1; -
DR	EMBL/ A009172; CA15657.1; -
DR	PIR; A35363; A35363.
DR	HSPB; P17599; IAUU.
DR	Gnew; HGNC:11494; SYNL.
DR	MIM; 313440; -
DR	GO; GO:0005803; C:secretory vesicle; TAS.
DR	GO; GO:0005215; F:transporter activity; TAS.
DR	GO; GO:0006899; P:membrane transport; TAS.
DR	GO; GO:0007268; P:synaptic transmission; TAS.
DR	InterPro; IPRO01359; Synapsin.
DR	Pfam; PF02078; Synapsin_1.
DR	Pfam; PF02750; Synapsin_C_1.
DR	PRINTS; PR01368; SYNAPSIN.
DR	PROSITE; PS00415; SYNAPSIN_1; 1.
DR	PROSITE; PS00416; SYNAPSIN_2; 1.
KW	Synapse; Phosphorylation; Neurone; Repeat; Actin-binding;
KW	Alternative splicing.
FT	DOMAIN 1 28
FT	DOMAIN 29 112
FT	DOMAIN 113 420
FT	A. B (LINKER).
FT	C
FT	(ACTIN-BINDING AND SYNAPTIC-VESSICLE BINDING)
FT	D (PRO-RICH LINKER).
FT	E. PHOSPHORYLATION (BY PKA AND CAMK1).
FT	F. PHOSPHORYLATION (BY CAMK2).
FT	G. MOD_RES 568 568
FT	MOD_RES 605 605
FT	VASAPPLIC 661 669
FT	/FtId-VSP 006316.
FT	/Missing (In Isoform Synapsin IB).
FT	/FtId-VSP 006317.
FT	G -> E (In REF. 3).
FT	SEQUENCE 705 AA; 73954 MW; 487831123PF6882P CRC64;
SO	
Query March	3.3%; Score 131; DB 1; Length 705;
Best Local Similarity 24.7%; Pred. No. 0.0024;	
Matches 76; Conservative 22; Mismatches 126; Indels 84; Gaps 13	
OY	209 VGSYHXAAEFRTKLVHPLIGNKSMDXVTVQVSFTLLMKCCCHGNPAQYEENRRKKXHL 268
DB	388 VGSSM-----PLIGHQDED---KQLVELVVVKMAQALPRGRDASP----- 428
OY	269 VYVLGXGANGAKIXSVGLXLMNAKSSEPGXTIPORRASVGLGXP----- 314
DB	429 ---GRSHQGTSPGALLPIGRGTSGQPAPRAQAQRPPQGSRPQFGPGRQGRPILQQ 484
OY	315 -----XXRLSPACRP-----PLHAPKXRGAARPPATGVRCAXVFVLNGAR 359
DB	485 PPPOGOOHLGLGPAGSPFLPQRLPEPTGAPQOPASQAPLPQGG-QRSHPVANGGAP 543
OY	360 PERPPXKRPSLTARSWGTXHT--XSISPPQGIKRS--PFIQOLTOPTHFG-----R 410
DB	544 PAAPRPASPS-PQRQAGRPRPATRGTSVGPARPKASGARPDGGQQRQGRPRQKRPBRAPTR 602
OY	411 EPAAKXSRPRADLP-AEERAPSKY-----PCLVQAEFEAVYEPKXQEETFYQPPVLQ 462
DB	603 QASQAGCVPTGCPETTCQRFPSGDPAGAPKQLAQKESDV-----PPPAT 649

463 QOXXGSEH 470  
650 AAGGPPH 657

Db

RESULT 13

ID MYTIF HUMAN STANDARD; FRT; 1098 AA.

AC 000160; Q6WMN7; 16-OCT-2001 (Rel. 40, Created)

DT 28-FEB-2003 (Rel. 41, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Myosin I $\beta$  (Myosin-IE).

GN MYOIF

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxId:9606;

RN [1]

RX MEDLINE=1664126; PubMed=1804589;

RA Krugman S., Anderson K.E., Ridley S.H., Rizzo N., McGregor A., Coadwell J., Davidson K., Eguinoa A., Elson C.D., Lipp P., Marileva M., Kisteleki N., Palmer G., Thuring J.H., Cooper M.A., Lim Z.-Y., Holmes A.B., Dove S.K., Michell R.H., Grewal A., Nasarain A., Erdjument-Bromage H., Tempst P., Stephens L.R., Hawkins P.T.;

RA "Identification of ARAP3, a novel PI3K effector regulating both Arf and Rho GTPases, by selective capture on phosphopeptide affinity matrices.";

RT Mol. Cell 9:95-108(2002).

RN [2]

RP SEQUENCE FROM N.A.

RC TISSUE=Splicein;

RP R1

RA Ishibashi T., Kanehori K., Yosida M., Matanabe S., Ishida S., Ono Y., Houta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J., Matanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y., Sugiyama T., Irie R., Otuki T., Sato H., Ota T., Makamatsu A., Ishii S., Yamamoto U., Isono Y., Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Matsuo K., Nakamura Y., Sekine M., Kikuchi H., Kanda K., Watanabe M., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B., Suzuki Y., Sugano S., Nagahara K., Maehuo Y., Nagai K., Isogai T.;

RA "NDO human cDNA sequencing project.";

RT Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.

RL [3]

RN SEQUENCE FROM N.A.

RC TISSUE=Blotot;

RC MEDLINE=22386257; PubMed=12477932;

RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Klausberg R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.F., Jordan H., Moore T., Max I., Wang J., Hsieh F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stempleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E., Brownstein M.J., Uedin T.B., Topolynski S., Carminci P., Pandey C., Rana S.S., McEwan P.U., McKernan K.J., Maier U.A., Gunaratne P.H., Bonak S.A., Wojcik P.C., Hale S., Garcia A.M., Gay L.J., Hultk S.W., Richards S., Wooley K.C., Hale S., Sodergren E.J., Lu X., Gibbs R.A., Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Pabey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield J.S.N., Krzywicki M.I., Skalska U., Smalls D.E., Schnerke A., Schein J.E., Jones S.U.W., Maita K.A.);

RT "Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [4]

RP SEQUENCE OF 318-1098 FROM N.A.

RC TISSUE=Retina;

RC MEDLINE=97237053; PubMed=9119401;

RA Crozet F., El Amraoui A., Blanchard S., Lenoir M., Ripoll C., Vago P.

RA Hamel C., Fizames C., Levi-Acobas F., Depetris D., Mattei M.-G.,  
 RA Weil D., Pujol R., Petit C.;  
 RT "Cloning of the genes encoding two murine and human cochlear  
 RT unconventional type I myosins";  
 RL Genomics 40:332-341(1997).  
 CC -I- FUNCTION: Myosins are actin-based motor molecules with ATPase  
 CC activity. Unconventional myosins serve in intracellular movements.  
 CC Their highly divergent tails are presumed to bind to membranous  
 CC compartments, which would be moved relative to actin filaments (By  
 CC similarity).  
 CC -I- SIMILARITY: BELONGS TO THE MYOSIN HEAVY CHAIN FAMILY.  
 CC -I- SIMILARITY: Contains 1 IQ domain.  
 CC -I- SIMILARITY: Contains 1 SH3 domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: AJ310570; CAC83948.1; -;  
 DR EMBL: AK092877; BAC03995.1; -;  
 DR EMBL: BC028071; AAH28071.1; -;  
 DR EMBL: X98411; CAA67058.1; -;  
 DR HSSP: P08799; 1MND.  
 DR Genew: HGNC:7600; MYO1F.  
 DR MIM: 601480; -;  
 DR GO: GO:0016461; C:unconventional myosin; NAS.  
 DR GO: GO:0003779; F:actin binding; NAS.  
 DR GO: GO:0005524; F:ATP binding; NAS.  
 DR GO: GO:0005516; F:calmodulin binding; NAS.  
 DR InterPro: IPR000048; IQ region.  
 DR InterPro: IPR001609; myosin\_head.  
 DR InterPro: IPR001452; SH3.  
 DR Pfam: PF00612; IQ; 1.  
 DR Pfam: PF00063; myosin\_head; 1.  
 DR Pfam: PF00018; SH3; 1.  
 DR PRINTS: PRO0193; MYOSINHEAVY.  
 DR PRINTS: PRO0452; SH3DOMAIN.  
 DR ProDom: PD000355; myosin\_head; 1.  
 DR ProDom: PD000066; SH3; 1.  
 DR SMART: SM00242; MYSC; 1.  
 DR SMART: SM00325; SH3; 1.  
 DR PROSITE: PS50096; IQ; 1.  
 DR PROSITE: PS50002; SH3; 1.  
 DR Myosin; ATP-binding; Actin-binding; Calmodulin-binding; SH3 domain;  
 KM Multigene family.  
 FT DOMAIN 1 677  
 FT DOMAIN 693 722  
 FT DOMAIN 1041 1088  
 FT NP\_BIND 579 589  
 FT NP\_BIND 110 117  
 FT NP\_BIND 259 259  
 FT NP\_BIND 534 536  
 FT NP\_BIND 592 593  
 FT NP\_BIND 602 602  
 FT NP\_BIND 759 759  
 FT NP\_BIND 797 797  
 FT NP\_BIND 805 805  
 FT NP\_BIND 809 809  
 FT NP\_BIND 814 814  
 FT NP\_BIND 922 922  
 FT NP\_BIND 927 927  
 FT NP\_BIND 930 930  
 FT NP\_BIND 948 948  
 FT NP\_BIND 958 958  
 FT NP\_BIND 971 971  
 FT NP\_BIND 1031 1031  
 FT NP\_BIND 1098 1098  
 FT NP\_BIND 1098 AA; 124803 MW; B7FCID371C421B8 CRC64;  
 SQ SEQUENCE

Query Match 3.2%; Score 129.5; DB 1; Length 1098;  
 Best Local Similarity 21.1%; Pred. No. 0.0052; Indels 103; Gaps 10;  
 Matches 58; Conservative 23; Mismatches 91;  
 QY 285 GLXINAKSKSRPGXTIRORRGASVGLGXPKXNLSPAPGPPUHAARXRGRARAPRAGT 344  
 DB 908 GRLTVTSVGGGLPKRSKPTKRG--MAKGPRRSSQAPTPAAPA-PPRGMDRNVPPSARG 964  
 QY 345 GVRXYPFVLGNARERPRRXPXPSRTASWGWRTYXGISSPQKLSPPFQXLTLP 404  
 DB 965 G---PLPLRTMSGGTHRPSPPS-----TSIGASRRPARPP-----SH 1003  
 QY 405 ETRF-----GREPAAXSRPADLPAPAPSPXPCLVQAEAEVYEE 447  
 DB 1004 NTEFLNVPDQMGAMGRKRSVGRFPVGVGRPK---PQRTHGPRC----- 1046  
 QY 448 PRQETFTYEQPLVVOQXXGSEHIDHIOGGLSXQGLCARALYDQADDXISPPEN 507  
 DB 1047 -----RALYQYVGQDVDELSPVNE 1066  
 QY 508 LITGIEVYXEGWNRG--YGPDGHFGMKXPANYVELI 540  
 DB 1067 VIELMEDEPSGWMKGRLLHGGELF---PGNYVEKI 1098  
 RESULT 14  
 NEBU HUMAN STANDARD; PRT; 6669 AA.  
 ID P20929; O15346;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE NEBULIN.  
 GN NEB.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID:9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95257391; PubMed=7739042;  
 RA Label S., Kolmerer B.;  
 RT "The complete primary structure of human nebulin and its correlation  
 RT to muscle structure";  
 RL J. Mol. Biol. 248:308-315(1995).  
 RN [2]  
 RP PARTIAL PRELIMINARY SEQUENCE.  
 RX MEDLINE=88284704; PubMed=3397062;  
 RA Zeviani M., Darras B.T., Rizzuto R., Salvati G., Betto R.,  
 RA Bonifati M., Miranda A.F., Du J., Samit C., Dickson G., Walsh F.S.,  
 RA Dimauro S., Francke U., Schon E.A.;  
 RT "Cloning and expression of human nebulin cDNAs and assignment of the  
 RT gene to chromosome 2q31-q32";  
 RL Genomics 2:249-256(1988).  
 RN [3]  
 RP STRUCTURE BY NMR OF 6610-6669.  
 RX MEDLINE=98179559; PubMed=9514727;  
 RA Poltoun A.S., Millevoi S., Gaudel M., Kolmerer B., Pastore A.;  
 RT "SH3 in muscles: solution structure of the SH3 domain from nebulin";  
 RL J. Mol. Biol. 276:189-202(1998).  
 CC -I- FUNCTION: This giant muscle protein may be involved in maintaining  
 CC the structural integrity of sarcomeres and the membrane system  
 CC associated with the myofibrils. Bind and stabilize F-actin.  
 CC -I- TISSUE SPECIFICITY: Muscle specific. Located in the thin filament  
 CC of striated muscle.  
 CC -I- DISEASE: Defects in NEB are a cause of the autosomal recessive  
 CC form of nemaline myopathy (NEM2).  
 CC -I- SIMILARITY: Contains 1 SH3 domain.  
 CC -I- SIMILARITY: Contains 178 nebulin repeats.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-stb.ch/announce/> or send an email to [license@isb-stb.ch](mailto:license@isb-stb.ch)).

-----

EMBL; X83957; CAA58788.1; ALT\_SEQ.  
EMBL; M19668; AAA59916.1; ALT\_SEQ.  
EMBL; M19669; AAA59917.1; ALT\_SEQ.  
PIR; A29979; A29979.  
PIR; B29979; B29979.  
PIR; S55024; S55024.  
PDB; INEB; 28-VAN-98.  
Gene; HGNC:7720; NEB.  
MIM; 161650; -.  
MIM; 256030; -.  
GO; GO:0015629; C:actin cytoskeleton; TAS.  
GO; GO:0030017; C:baroreceptor; NAS.  
GO; GO:0008307; P:structural constituent of muscle; TAS.  
GO; GO:0030832; P:regulation of actin filament length; NAS.  
GO; GO:0007525; P:somatic muscle development; NAS.  
InterPro; IPR000900; Nebulin.  
InterPro; IPR001452; SH3.  
Pfam; PF00880; Nebulin; 146.  
Pfam; PF00018; SH3; 1.  
PRINTS; PRO0510; NEBULIN.  
PRINTS; PRO0452; SH3DOMAIN.  
ProDom; PD000066; SH3; 1.  
ProSITE; PS50002; SH3; 1.  
Actin-binding; Muscle protein; Cytoskeleton; Repeat; 3D-structure.

REPEAT 76 107 NEBULIN 1.  
REPEAT 113 143 NEBULIN 2.  
REPEAT 148 178 NEBULIN 3.  
REPEAT 183 213 NEBULIN 4.  
REPEAT 218 248 NEBULIN 5.  
REPEAT 253 283 NEBULIN 6.  
REPEAT 289 318 NEBULIN 7.  
REPEAT 324 354 NEBULIN 8.  
REPEAT 363 393 NEBULIN 9.  
REPEAT 398 428 NEBULIN 10.  
REPEAT 434 464 NEBULIN 11.  
REPEAT 502 532 NEBULIN 12.  
REPEAT 537 567 NEBULIN 13.  
REPEAT 573 603 NEBULIN 14.  
REPEAT 611 641 NEBULIN 15.  
REPEAT 681 711 NEBULIN 16.  
REPEAT 749 779 NEBULIN 17.  
REPEAT 814 844 NEBULIN 18.  
REPEAT 820 850 NEBULIN 19.  
REPEAT 858 888 NEBULIN 20.  
REPEAT 893 923 NEBULIN 21.  
REPEAT 924 954 NEBULIN 22.  
REPEAT 959 990 NEBULIN 23.  
REPEAT 993 1023 NEBULIN 24.  
REPEAT 1028 1058 NEBULIN 25.  
REPEAT 1064 1094 NEBULIN 26.  
REPEAT 1102 1132 NEBULIN 27.  
REPEAT 1137 1167 NEBULIN 28.  
REPEAT 1168 1198 NEBULIN 29.  
REPEAT 1204 1234 NEBULIN 30.  
REPEAT 1237 1267 NEBULIN 31.  
REPEAT 1272 1302 NEBULIN 32.  
REPEAT 1308 1338 NEBULIN 33.  
REPEAT 1346 1376 NEBULIN 34.  
REPEAT 1381 1411 NEBULIN 35.  
REPEAT 1412 1442 NEBULIN 36.  
REPEAT 1448 1478 NEBULIN 37.  
REPEAT 1481 1511 NEBULIN 38.  
REPEAT 1516 1546 NEBULIN 39.  
REPEAT 1552 1582 NEBULIN 40.  
REPEAT 1590 1620 NEBULIN 41.  
REPEAT 1625 1655 NEBULIN 42.

REPEAT 1656 1686 NEBULIN 43.  
REPEAT 1692 1722 NEBULIN 44.  
REPEAT 1725 1755 NEBULIN 45.  
REPEAT 1760 1790 NEBULIN 46.  
REPEAT 1796 1826 NEBULIN 47.  
REPEAT 1834 1864 NEBULIN 48.  
REPEAT 1869 1899 NEBULIN 49.  
REPEAT 1900 1930 NEBULIN 50.  
REPEAT 1936 1966 NEBULIN 51.  
REPEAT 1969 1999 NEBULIN 52.  
REPEAT 2004 2034 NEBULIN 53.  
REPEAT 2040 2070 NEBULIN 54.  
REPEAT 2078 2108 NEBULIN 55.  
REPEAT 2113 2143 NEBULIN 56.  
REPEAT 2144 2174 NEBULIN 57.  
REPEAT 2180 2210 NEBULIN 58.  
REPEAT 2213 2243 NEBULIN 59.  
REPEAT 2248 2278 NEBULIN 60.  
REPEAT 2284 2314 NEBULIN 61.  
REPEAT 2322 2352 NEBULIN 62.  
REPEAT 2357 2387 NEBULIN 63.  
REPEAT 2388 2418 NEBULIN 64.  
REPEAT 2423 2453 NEBULIN 65.  
REPEAT 2456 2486 NEBULIN 66.  
REPEAT 2491 2521 NEBULIN 67.  
REPEAT 2527 2557 NEBULIN 68.  
REPEAT 2565 2595 NEBULIN 69.  
REPEAT 2600 2630 NEBULIN 70.  
REPEAT 2631 2661 NEBULIN 71.  
REPEAT 2666 2696 NEBULIN 72.  
REPEAT 2699 2729 NEBULIN 73.  
REPEAT 2734 2764 NEBULIN 74.  
REPEAT 2770 2800 NEBULIN 75.  
REPEAT 2808 2838 NEBULIN 76.  
REPEAT 2843 2873 NEBULIN 77.  
REPEAT 2874 2904 NEBULIN 78.  
REPEAT 2909 2939 NEBULIN 79.  
REPEAT 2942 2972 NEBULIN 80.  
REPEAT 2977 3007 NEBULIN 81.  
REPEAT 3013 3043 NEBULIN 82.  
REPEAT 3051 3081 NEBULIN 83.  
REPEAT 3086 3116 NEBULIN 84.  
REPEAT 3117 3147 NEBULIN 85.  
REPEAT 3152 3182 NEBULIN 86.  
REPEAT 3185 3215 NEBULIN 87.  
REPEAT 3220 3250 NEBULIN 88.  
REPEAT 3256 3286 NEBULIN 89.  
REPEAT 3294 3324 NEBULIN 90.  
REPEAT 3329 3359 NEBULIN 91.  
REPEAT 3360 3390 NEBULIN 92.  
REPEAT 3395 3425 NEBULIN 93.  
REPEAT 3428 3458 NEBULIN 94.  
REPEAT 3463 3493 NEBULIN 95.  
REPEAT 3499 3529 NEBULIN 96.  
REPEAT 3537 3567 NEBULIN 97.  
REPEAT 3572 3602 NEBULIN 98.  
REPEAT 3603 3633 NEBULIN 99.  
REPEAT 3638 3668 NEBULIN 100.  
REPEAT 3671 3701 NEBULIN 101.  
REPEAT 3706 3736 NEBULIN 102.  
REPEAT 3742 3772 NEBULIN 103.  
REPEAT 3780 3810 NEBULIN 104.  
REPEAT 3815 3845 NEBULIN 105.  
REPEAT 3846 3876 NEBULIN 106.  
REPEAT 3914 3944 NEBULIN 107.  
REPEAT 3949 3979 NEBULIN 108.  
REPEAT 3984 4014 NEBULIN 109.  
REPEAT 4021 4052 NEBULIN 110.  
REPEAT 4057 4087 NEBULIN 111.  
REPEAT 4088 4118 NEBULIN 112.  
REPEAT 4123 4153 NEBULIN 113.  
REPEAT 4156 4186 NEBULIN 114.  
REPEAT 4191 4220 NEBULIN 115.

```

FT REPEAT 4226 4256 NEBULIN 116.
FT REPEAT 4264 4294 NEBULIN 117.
FT REPEAT 4299 4329 NEBULIN 118.
FT REPEAT 4330 4360 NEBULIN 119.
FT REPEAT 4365 4395 NEBULIN 120.
FT REPEAT 4400 4430 NEBULIN 121.
FT REPEAT 4435 4465 NEBULIN 122.
FT REPEAT 4471 4501 NEBULIN 123.
FT REPEAT 4544 4574 NEBULIN 124.

Query Match 3.2%; Score 126; DB 1; Length 6669;
Best Local Similarity 27.0%; Pred. No. 0.069;
Matches 47; Conservative 13; Mismatches 72; Indels 42; Gaps 5;

QY 402 TQPTHTFRPAPAXSRP-----ADLPAEPAPSPPCLYQAEF-----442
DB 6503 TNPQSVFDPYDAEDNIQGRSLHMTNVQQRSRQSSASALSVSGGEKSEHSEADPH 6562
QY 443 -----AYEEFEQETFEYEQPPLVQOQXGSEHIDHIIQGGLSXQGLCARA 489
DB 6563 LSTVSDGCVFAVSTAYKAKTELPQQRSSSVATQCTVSSIPSH-----PSTAGKIFRA 6617
QY 450 LYDYQADDXEISPDENLITGIEVXXEGMWRGYP---DGHFGKXPANYELI 540
DB 6618 MYDYMAADADRVSPFDGDAIINVQAIQDEGMW--YGTVQRTGRTMLPANYEAL 6669

RESULT 15
ID ITN1 RAT STANDARD; PRT; 1217 AA.
AC Q9WVE9; Q9WVE1;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Intersectin 1 (EH domain and SH3 domain regulator of endocytosis 1).
GN ITSN1 OR ITSN OR EHSN1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Brain;
RX MEDLINE=99303609; PubMed=10373452;
RA Okamoto M., Schoch S., Suehof T.C.;
RT "EHSN1/Intersectin, a protein that contains EH and SH3 domains and
RT binds to dynamin and SNAP-25. A protein connection between exocytosis
RT and endocytosis?";
RL J. Biol. Chem. 274:18446-18454 (1999).
CC -1- FUNCTION: Adapter protein that may provide indirect link between
CC the endocytic membrane traffic and the actin assembly machinery.
CC May regulate the formation of clathrin-coated vesicles.
CC -1- SUBUNIT: Interacts with dynamin, SNAP-25 and SNAP-23. Clusters
CC several dynamin in a manner that is regulated by alternative
CC splicing. Also binds clathrin-associated proteins and other
CC components of the endocytic machinery, such as N-WASP, Epsins and
CC Stomatin 2 (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic; membrane-associated protein.
CC -1- Enriched in synaptosomes.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=Q9WVE9-1; Sequence=Displayed;
CC Name=2;
CC IsoId=Q9WVE9-2; Sequence=VSP_004297;
CC -1- TISSUE SPECIFICITY: Highly expressed in brain.
CC -1- DOMAIN: SH3-3, SH3-4 and SH3-5, but not SH3-1 and SH3-2 domains,
CC bind to dynamin.
CC -1- DOMAIN: The KLRQ domain binds to SNAP-25 and SNAP-23.
CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
CC -1- SIMILARITY: Contains 5 SH3 domains.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AF127798; AAD30271.1; -
CC EMBL; AF132672; AAD31026.1; -
CC HSSP; P29354; 1GFC.
CC InterPro: IPR002048; EF-hand.
CC InterPro: IPR000261; Eps15_homology.
CC InterPro: IPR001452; SH3.
CC Pfam; PF00036; ehnd; 3.
CC Pfam; PF00018; SH3; 5.
CC PRINTS; PR00452; SH3DOMAIN.
CC ProDom; PD000066; SH3; 5.
CC SMART; SM00054; Eph; 2.
CC SMART; SM00027; EH; 2.
CC SMART; SM00326; SH3; 5.
CC PROSITE; PS00018; EF_HAND; 2.
CC PROSITE; PS00031; EH; 2.
CC PROSITE; PS00002; SH3; 5.
CC KEGG; Endocytosis; SH3 domain; Repeat; Coiled coil; Calcium-binding;
CC Alternative splicing.
CC FT DOMAIN 21 109
CC FT CA_BIND 66 78 EH 1.
CC FT DOMAIN 221 310 EF_HAND 1 (POTENTIAL).
CC FT CA_BIND 267 279 EH 2.
CC FT DOMAIN 326 702 LYS/LEU/GLU/ARG/GLN-RICH (KERO).
CC FT DOMAIN 350 670 LYS/LEU/GLU/ARG/GLN-RICH (KERO).
CC FT DOMAIN 737 803 COILED COIL (POTENTIAL).
CC FT DOMAIN 910 968 SH3 1.
CC FT DOMAIN 968 999 SH3 2.
CC FT DOMAIN 999 1057 SH3 3.
CC FT DOMAIN 1071 1135 SH3 4.
CC FT DOMAIN 1152 1211 SH3 5.
CC FT DOMAIN 1321 134 POLY-SER.
CC FT VARSPDIC 1003 1073 Missing (in isoform 2).
CC FT /FTId=VSP_004297.
CC SQ SEQUENCE 1217 AA; 137154 MW; 6C13238A5A5B34B CRC64;

Query Match 3.1%; Score 124.5; DB 1; Length 1217;
Best Local Similarity 23.0%; Pred. No. 0.015;
Matches 41; Conservative 19; Mismatches 47; Indels 71; Gaps 5;

QY 368 PSRT-----ARSGTWKXTXSSPOPKLRSPFLQXQLTQPTHTFRPAPAXSRPAD 422
DB 855 PSTNKEPRTDWDITRAQPSLTIVPSAGLR-----QRAFTPATATGSSRSP-----902
QY 423 LPAEPAPSPPCLYQAEZEAAYEPEQETFEYEQPPLVQOQXGSEHIDHIIQGGLSX 482
DB 903 -----VLQGGKRV 910
QY 483 QGLCARALYDYQADDXEISPDENLITGIEVXXEGMWRGYPDGHFGKXPANYELI 540
DB 911 EQLQALYPRPAKXNDHNLNFKKSVITVLB-QQDMWFG-EVQQKGMFRRSYVKLI 966

```

Search completed: April 27, 2004, 10:53:37  
Job time : 16.5828 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OK protein - protein search, using sw model

Run on: April 27, 2004, 10:45:34 ; Search time 42.2397 Seconds  
(without alignments)  
5923.479 Million cell updates/sec

Title: US-10-028-952A-9

Perfect score: 3989  
Sequence: 1 HEIPVTPVRYPAKPOXHERA.....DFKSKKKKKLTKLMTSSX 793

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	796	20.0	302	4	Q9HBN8
2	796	20.0	430	4	Q9JUB6
3	796	20.0	430	4	Q9NR72
4	796	20.0	431	4	Q9EF30
5	794	19.9	439	4	Q96K74
6	628.5	15.8	433	11	Q6C2418
7	628.5	15.8	436	11	Q60WPI
8	627.5	15.7	432	11	Q8BH56
9	606	15.2	429	11	Q9JW74
10	603.5	15.1	432	11	Q9JW66
11	602	15.1	433	11	Q9JW67
12	599.5	15.0	436	11	Q9JH14
13	365.5	9.2	447	13	Q7ZXC9
14	231	5.8	559	5	Q9YDF4
15	228.5	5.7	486	11	Q9Z218
16	227.5	5.7	384	5	Q8MRS9

17	227.5	5.7	531	5	Q9VUB4	Q9VUB4 drosophila
18	226.5	5.7	643	5	Q9XUT0	Q9XUT0 caenorhabd
19	223	5.6	559	5	Q96046	Q96046 drosophila
20	210	5.3	530	13	Q8UWC3	Q8UWC3 xenopus lae
21	209	5.2	471	11	Q70419	Q70419 rattus norv
22	209	5.2	508	11	Q70420	Q70420 rattus norv
23	208	5.2	509	11	Q92116	Q92116 mus musculu
24	208	5.2	509	11	Q8BNAS	Q8BNAS mus musculu
25	207.5	5.2	477	5	Q96E20	Q96E20 suberites d
26	205	5.1	513	4	Q96H99	Q96H99 homo sapien
27	201	5.0	587	5	Q96459	Q96459 strongyloce
28	184.5	4.6	206	13	Q7SXH4	Q7SXH4 brachydantio
29	168	4.2	758	5	Q6BK16	Q6BK16 dictyostell
30	167.5	4.2	634	4	Q8N707	Q8N707 homo sapien
31	160.5	4.0	513	4	Q9NWB9	Q9NWB9 homo sapien
32	153.5	3.8	239	4	Q13802	Q13802 homo sapien
33	152	3.8	475	4	Q13801	Q13801 homo sapien
34	148.5	3.7	390	4	Q13249	Q13249 homo sapien
35	148	3.7	401	4	Q13147	Q13147 homo sapien
36	147	3.7	543	11	Q35823	Q35823 rattus norv
37	146	3.7	263	11	Q99WZ8	Q99WZ8 rattus norv
38	146	3.7	368	3	Q9P837	Q9P837 candida alb
39	146	3.7	475	4	Q9BV70	Q9BV70 homo sapien
40	145	3.6	263	6	Q77506	Q77506 oryctolagus
41	143	3.6	315	13	Q03292	Q03292 xenopus lae
42	140.5	3.5	473	5	Q9Y0S9	Q9Y0S9 drosophila
43	136.5	3.4	447	5	Q86K04	Q86K04 dictyostell
44	134.5	3.4	1098	11	Q8CG29	Q8CG29 mus musculu
45	134.5	3.4	1098	11	Q811E7	Q811E7 mus musculu

## ALIGNMENTS

## RESULT 1

Q9HBN8 PRELIMINARY; PRT; 302 AA.

AC Q9HBN8; 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DE 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
DS Hypothetical protein.  
OS Homo sapiens (human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Gu J.R., Man D.F., Zhao X.T., Zhou X.M., Jiang H.Q., Zhang P.P.,  
RA Qin W.X., Huang Y., Qiu X.K., Qian L.F., He L.P., Li H.N., Yu Y.,  
RA Yu J., Han L.H.,  
RT "Novel Human cDNA clones with function of inhibiting cancer cell  
RT growth."  
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
DR EMBL; AF218020; AAG17262.1; --  
DR HSSP; P23727; 1PNU.  
DR InterPro; IPR001452; SH3.  
DR Pfam; PF00018; SH3; 1.  
DR PRINTS; PR00452; SH3DOMAIN.  
DR PRODOM; PD000066; SH3; 1.  
DR SMART; SMO0326; SH3; 1.  
DR PROSITE; PS50002; SH3; 1.  
KM Hypothetical protein; SH3 domain.  
SQ SEQUENCE 302 AA; 34388 MW; 2DECA14C7511612 CRC64;

Query Match 20.0%; Score 796; DB 4; Length 302;  
Best Local Similarity 82.5%; Pred. No. 1.1e-70;  
Matches 151; Conservative 2; Mismatches 30; Indels 0; Gaps 0;

QY 359 RPERKXRRSRTARSGWTRXTSISPPGKLRSPFLQXOTOPETGGRPAAXSR 418  
DB 120 RNEQSAVHREIRFKQERAMSTTSISPPGKLRSPFLQKULTOPETHGRPAALSR 179

QY 419 PRADIPAEPPAPSPPCLVQAEAEVVEEPKQETFEYQPPVVOOQXXGSEHIDHIQOQ 478  
 DB 180 PRADIPAEPPAPSPPCLVQAEAEVVEEPKQETFEYQPPVVOOQXXGSEHIDHIQOQ 239  
 QY 479 GLSXQGLCARALYDYQAAADDEISFPDENLITGIEVXXEGWWRGYPDGHFGKMPANVYE 538  
 DB 240 GLSXQGLCARALYDYQAAADDEISFPDENLITGIEVDEGWRGYPDGHFGKMPANVYE 299  
 QY 539 LIE 541  
 DB 300 LIE 302

RESULT 2  
 ID Q9UT06 PRELIMINARY; PRT; 430 AA.  
 AC Q9UT06;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE SRC homolog 3 domain-containing protein HIP-55 (Drebrin F).  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20036526; PubMed=10567356;  
 RA Eisenat D., Yao Z., Wang X.-S., Kori R., Zhou G., Lee S.C., Tan T.-H.,  
 RT "A Novel Src Homology 3 Domain-containing Adaptor Protein, HIP-55,  
 RT that Interacts with Hematopoietic Progenitor Kinase 1.";  
 RL J. Biol. Chem. 274:33945-33950(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Zhang W., Yuan Z., Wan T., He L., Gao X.;  
 RT "Molecular cloning of cDNA encoding drebrin F";  
 RN Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
 RL [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Straubeberg R.;  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF197060; AAT3701.1; -;  
 DR EMBL; AF077353; AAF80228.1; -;  
 DR EMBL; BC031687; AAB31687.1; -;  
 DR HSSP; P23727; IPNU.  
 DR GO; GO:0008047; F:enzyme activator activity; TAS.  
 DR GO; GO:0005515; F:protein binding; TAS.  
 DR GO; GO:0007257; F:activation of JUNK; TAS.  
 DR InterPro; IPR00108; Actbind\_cofin.  
 DR InterPro; IPR001452; SH3.  
 DR Pfam; PF00241; Cofilin\_ADP; 1.  
 DR Pfam; PF00018; SH3; 1.  
 DR PRINTS; PR00452; SH3DOMAIN.  
 DR Prodom; PD000066; SH3; 1.  
 DR PROSITE; PSS0002; SH3; 1.  
 SQ SEQUENCE 430 AA; 48207 MW; 7B8C42ED047257AE CRC64;

Query Match 20.0%; Score 796; DB 4; Length 430;  
 Best Local Similarity 82.5%; Pred. No. 1.6e-70;  
 Matches 151; Conservative 2; Mismatches 30; Indels 0; Gaps 0;

QY 359 RPERPXXRPSRTARSGTWGTXYSISSPOGKLRSPFLQOLQPEFHFRERPAAXSR 418  
 DB 248 RNEBSAVHPRERIKOKERAMSTTSSISPOGKLRSPFLQOLQPEFHFRERPAAXSR 307  
 QY 419 PRADIPAEPPAPSPPCLVQAEAEVVEEPKQETFEYQPPVVOOQXXGSEHIDHIQOQ 478  
 DB 308 PRADIPAEPPAPSPPCLVQAEAEVVEEPKQETFEYQPPVVOOQXXGSEHIDHIQOQ 367  
 QY 479 GLSXQGLCARALYDYQAAADDEISFPDENLITGIEVXXEGWWRGYPDGHFGKMPANVYE 538

DB 368 GLSXQGLCARALYDYQAAADDEISFPDENLITGIEVDEGWRGYPDGHFGKMPANVYE 427  
 QY 539 LIE 541  
 DB 428 LIE 430

RESULT 3  
 ID Q9NR72 PRELIMINARY; PRT; 430 AA.  
 AC Q9NR72;  
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Cervical SH3p7 (Mucin-associated protein).  
 OS Homo sapiens (human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ho S.B., Toribara N.W., Anway R.E., Spurr-Michaud S.J., Shekels L.L.,  
 RA Keutmann H.T., Hill J.A., Gipson I.K.;  
 RT "Expression cloning of human cervical proteins using an antibody to  
 RT cervical mucus";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Cervix;  
 RX MEDLINE=99077625; PubMed=9858486;  
 RA Gipson I.K., Spurr-Michaud S., Moccia R., Zhan O., Toribara N.,  
 RA Ho S.B., Garduno A.R., Hill J.A., Iii;  
 RT "hNC4 and MUC5B transcripts are the prevalent mucin messenger  
 RT ribonucleic acids of the human endocervix";  
 RL Biol. Reprod. 60:58-64(1999).  
 CC -1 SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 DR EMBL; AF250287; AAF81273.1; -;  
 DR EMBL; AF151364; AAG3120.1; -;  
 DR HSSP; P23727; IPNU.  
 DR GO; GO:0005622; C:intracellular; IEA.  
 DR GO; GO:0003779; F:actin binding; IEA.  
 DR InterPro; IPR002108; Actbind\_cofin.  
 DR InterPro; IPR001452; SH3.  
 DR Pfam; PF00241; Cofilin\_ADP; 1.  
 DR Pfam; PF00018; SH3; 1.  
 DR PRINTS; PR00452; SH3DOMAIN.  
 DR Prodom; PD000066; SH3; 1.  
 DR SMART; SM00102; ADP; 1.  
 DR SMART; SM00326; SH3; 1.  
 DR PROSITE; PSS0002; SH3; 1.  
 KV SH3 domain.  
 SQ SEQUENCE 430 AA; 48154 MW; 7D0CAF17D8CEB8DC CRC64;

Query Match 20.0%; Score 796; DB 4; Length 430;  
 Best Local Similarity 82.5%; Pred. No. 1.6e-70;  
 Matches 151; Conservative 2; Mismatches 30; Indels 0; Gaps 0;

QY 359 RPERPXXRPSRTARSGTWGTXYSISSPOGKLRSPFLQOLQPEFHFRERPAAXSR 418  
 DB 248 RNEBSAVHPRERIKOKERAMSTTSSISPOGKLRSPFLQOLQPEFHFRERPAAXSR 307  
 QY 419 PRADIPAEPPAPSPPCLVQAEAEVVEEPKQETFEYQPPVVOOQXXGSEHIDHIQOQ 478  
 DB 308 PRADIPAEPPAPSPPCLVQAEAEVVEEPKQETFEYQPPVVOOQXXGSEHIDHIQOQ 367  
 QY 479 GLSXQGLCARALYDYQAAADDEISFPDENLITGIEVXXEGWWRGYPDGHFGKMPANVYE 538  
 DB 368 GLSXQGLCARALYDYQAAADDEISFPDENLITGIEVDEGWRGYPDGHFGKMPANVYE 427  
 QY 539 LIE 541  
 DB 428 LIE 430



RESULT 4  
ID 096F30 PRELIMINARY; PRT; 431 AA.  
AC 096F30  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Similar to src homology 3 domain-containing protein HRP-55.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
NCBI\_TaxID=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE=Eye;  
RA Strausberg R.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
CC EMBL; BC011677; A011677.1; -  
DR GO; GO:0005622; C:intracellular; IEA.  
DR GO; GO:0003779; F:actin binding; IEA.  
DR InterPro; IPR002108; F:actin\_bind\_cofln.  
DR InterPro; IPR001452; SH3.  
DR Pfam; PF00241; coflin\_ADF; 1.  
DR Pfam; PF00018; SH3; 1.  
DR PRINTS; PR00452; SH3DOMAIN.  
DR PRODOM; PD000066; SH3; 1.  
DR SMART; SM00102; ADF; 1.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PS50002; SH3; 1.  
KW SH3 domain.  
SQ SEQUENCE 431 AA; 48294 MW; D18P9D316FFD0B4E CRC64;

Query Match 20.0%; Score 796; DB 4; Length 431;  
Best Local Similarity 82.5%; Pred. No. 1.6e-70;  
Matches 151; Conservative 2; Mismatches 30; Indels 0; Gaps 0;

QY 359 REPRPPXKPRSTARSWGTWRTXKSISSPQPKLSPFQXQLOLTPETHFGREPAAXSR 418  
DB 248 RVEQSAVPRFETFKOKERAMSTSISSPQPKLSPFQXQLOLTPETHFGREPAAXSR 308  
QY 419 PRADLPAAEPAPSPPCIVQAEAEAYEEPEKQETFFYQPLVQOQXXGSEHIDHIOGQ 478  
DB 309 PRADLPAAEPAPSPPCIVQAEAEAYEEPEKQETFFYQPLVQOQXXGSEHIDHIOGQ 368  
QY 479 GTSXQGLCARALYDQADDXEISPDPELITGIEVXXEGMWRGYPDGHFGMPANYVE 538  
DB 369 GTSXQGLCARALYDQADDXEISPDPELITGIEVXXEGMWRGYPDGHFGMPANYVE 428  
QY 539 LIE 541  
DB 429 LIE 431

RESULT 5  
ID 096K74 PRELIMINARY; PRT; 439 AA.  
AC 096K74  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Hypothetical protein FL114461.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
NCBI\_TaxID=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE=Mammary gland;  
RA Isegai T., Ota T., Hayaashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,

RA Nagatsuna M., Hosoiri T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
RA Takahashi K., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,  
RA Matanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Masuno Y.,  
RA Niimura K., Iwayanagi T.,  
RT "NEDO human cDNA sequencing project."  
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
DR EMBL; AK027367; BA85065.1; -  
DR GO; GO:0005622; C:intracellular; IEA.  
DR GO; GO:0003779; F:actin binding; IEA.  
DR InterPro; IPR002108; F:actin\_bind\_cofln.  
DR InterPro; IPR001452; SH3.  
DR Pfam; PF00241; coflin\_ADF; 1.  
DR Pfam; PF00018; SH3; 1.  
DR PRINTS; PR00452; SH3DOMAIN.  
DR PRODOM; PD000066; SH3; 1.  
DR SMART; SM00102; ADF; 1.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PS50002; SH3; 1.  
KW Hypothetical protein; SH3 domain.  
SQ SEQUENCE 439 AA; 49042 MW; 8C044FBD0B82C6D5 CRC64;

Query Match 19.3%; Score 794; DB 4; Length 439;  
Best Local Similarity 91.3%; Pred. No. 2.6e-70;  
Matches 148; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 381 TSISSPQPKLSPFQXQLOLTPETHFGREPAAXSRPRADLPAAEPAPSPPCIVQAE 440  
DB 279 TSISSPQPKLSPFQXQLOLTPETHFGREPAAXSRPRADLPAAEPAPSPPCIVQAE 338  
QY 441 EBAVYEEPEKQETFFYQPLVQOQXXGSEHIDHIOGQGLCARALYDQADDXE 500  
DB 339 EBAVYEEPEKQETFFYQPLVQOQXXGSEHIDHIOGQGLCARALYDQADDXE 398  
QY 501 ISPDPELITGIEVXXEGMWRGYPDGHFGMPANYVE 541  
DB 399 ISPDPELITGIEVXXEGMWRGYPDGHFGMPANYVE 439

RESULT 6  
ID 062418 PRELIMINARY; PRT; 433 AA.  
AC 062418  
DT 01-JUN-1998 (TREMBlrel. 06, Created)  
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE Drebilin-like SH3 domain-containing protein SH3P7.  
GN DBIL.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxID=10090;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE=EMBRYO;  
RX MEDLINE=98234438; Pubmed=9630982;  
RA Sparks A.B., Hoffman N.G., McConnell S.J., Fowlkes D.M., Kay B.K.,  
RL Nat. Biotechnol. 14:741-744(1996).  
CC -1- SIMILARITY: CONTAINS A COPY OF THE SH3 DOMAIN.  
CC -1- SIMILARITY: TO DREBRIN.  
DR EMBL; U58884; AAC52640.1; -  
DR HSP; P19174; ZHSP.  
DR MGI; MGI:700006; Dbnl.  
DR GO; GO:0005737; C:cytosol; IEA.  
DR GO; GO:0030027; C:lamellipodium; IEA.  
DR GO; GO:0003779; F:actin binding; IEA.  
DR GO; GO:0005515; F:protein binding; IEA.  
DR GO; GO:0016601; P:RAC protein signal transduction; IEA.  
DR InterPro; IPR002108; F:actin\_bind\_cofln.  
DR InterPro; IPR001452; SH3.  
DR Pfam; PF00241; coflin\_ADF; 1.  
DR Pfam; PF00018; SH3; 1.

PRINTS: PR00452; SHDOMAIN.  
 DR PRODOM: PD00006; SH3; 1.  
 DR SMART; SMO0102; ADF; 1.  
 DR SMART; SMO0326; SH3; 1.  
 DR PROSITE; PS00002; SH3; 1.  
 KW SH3 domain.  
 FT DOMAIN 356 432 SH3.  
 SQ SEQUENCE 433 AA; 48428 MW; 602D3B62C446FAAD CRC64;

Query Match 15.8%; Score 628.5; DB 11; Length 433;  
 Best Local Similarity 58.7%; Pred. No. 7.7e-54;  
 Matches 131; Conservative 12; Mismatches 43; Indels 37; Gaps 5;

QY 336 AAPPPAGTGVRCXVFLVNGARDEPRPXXRPSRTASWGT-----W 378  
 DB 231 AGAPSPRTG-----EPECEAV---SRTQEWESAGQAHPREIFKOKERA 272  
 QY 379 RXTXSISSPQPKLSPFLQXLTQPEETHFGREPAAXSRPADLPAREPAPSPCTLVQ 438  
 DB 273 KSTTSVTSQPKLSPFLQXLTQPEETHFGREPAAXSRPADLPAREPAPSPCTLVQ 330  
 QY 439 AEEBAVEEPXQETFFYQPPVQOQXXGSHIDHIIQGGSLXQGLCAPALYDYQAAD 498  
 DB 331 TEEEPYVEPPEODTLVEEPPLVQOQAGSEHIDHIIQGGSLXQGLCAPALYDYQAAD 390  
 QY 499 XEISPDPEPLTIGIEVXXEGMWRGYPDGHFGKXPANVVELI 541  
 DB 391 TEISPDPEPLTIGIEVDEGMWRGYPDGHFGKXPANVVELI 433

RESULT 7  
 ID 080WP1 PRELIMINARY; PRT; 436 AA.  
 AC 080WP1;  
 DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Actin-binding protein 1.  
 GN ABP1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL6/J;  
 RA Hou P., Estrada L., Gorecki J.D.,  
 RT "Fgdl, a Cdc42-specific guanine nucleotide exchange factor, directly  
 RT interacts with cortactin and Ablp1 to modulate cell shape.",  
 RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AY098595; AA028340.1;  
 DR GO; GO:0005622; C:intracellular; IEA.  
 DR GO; GO:0003779; F:actin binding; IEA.  
 DR InterPro; IPR002108; Actbind\_cofan.  
 DR InterPro; IPR001452; SH3.  
 DR Pfam; PF00241; Cofilin\_ADF; 1.  
 DR Pfam; PF00018; SH3; 1.  
 DR PRINTS; PR00452; SH3DOMAIN.  
 DR PRODOM; PD00006; SH3; 1.  
 DR SMART; SMO0102; ADF; 1.  
 DR SMART; SMO0326; SH3; 1.  
 DR PROSITE; PS00002; SH3; 1.  
 DR PROSITE; PS00002; SH3; 1.  
 SQ SEQUENCE 436 AA; 48699 MW; 85AEF9781C698A3F CRC64;

Query Match 15.8%; Score 628.5; DB 11; Length 436;  
 Best Local Similarity 55.2%; Pred. No. 7.8e-54;  
 Matches 133; Conservative 15; Mismatches 58; Indels 35; Gaps 5;  
 QY 318 LSPPRGPPPLAPRKGARAPPAAGTCVRCXVFLVNGARDEPRPXXRPSRTASWGT 377  
 DB 214 LOEARRRORYOQHRSGAPSPSSRTG-----EPQCEAV---SRTQEWES 257  
 QY 378 -----WRXTXSISSPQPKLSPFLQXLTQPEETHFGREPAAXSRPR 420

DB 258 AGQAHPREIFKOKERAMSTTSVTSQPKLSPFLQXLTQPEETHFGREPAAXSRRA 317  
 QY 421 ADLPAEPAPSPCTLVQAEBAVEEPXQETFFYQPPVQOQXXGSHIDHIIQGGSL 480  
 DB 318 AGV-CEEPAPSTLSS-AQTEEPYVEPPEODTLVEEPPLVQOQAGSEHIDHIIQGGSL 375  
 QY 481 SXQGLCAPALYDYQAADDEXISPDPEPLTIGIEVXXEGMWRGYPDGHFGKXPANVVELI 540  
 DB 376 SGQGLCAPALYDYQAADDTISPDPEPLTIGIEVDEGMWRGYPDGHFGKXPANVVELI 435  
 QY 541 E 541  
 DB 436 E 436

RESULT 8  
 ID 08BH56 PRELIMINARY; PRT; 432 AA.  
 AC 08BH56;  
 DT 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Drebrin-1 like.  
 GN BDNL.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Brain, Eye, and Medulla oblongata;  
 RX MEDLINE=22354683; PubMed=12466851;  
 RA The FANTOM Consortium.  
 RA "The RIKEN Genome Exploration Research Group Phase I & II Team,  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.",  
 RL Nature 420:563-573 (2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Colon;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Sherman C.M., Schler G.D.,  
 RA Altschul S.F., Zeeberg B., Bueltow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stopleton M.J., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Millar S.J.,  
 RA Baha S.S., Loguclano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McKernan P.J., McKernan K.J., Malek U.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalón D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs S.A.,  
 RA Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Greenwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smalhus D.E., Scherch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences",  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Colon;  
 RA Straubeberg R.;  
 RL Submitted (Feb-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AK046073; BAC32592.1; -  
 DR EMBL; AK053796; BAC35528.1; -  
 DR EMBL; AK078082; BAC37118.1; -  
 DR EMBL; BC046430; AAH46430.1; -  
 DR MGD; MGI:700006; Dbnl.  
 DR GO; GO:0005737; C:Cytoplasm; IDA.

DR GO:0030027; C:lamellipodium; IDA.  
DR GO:0003779; F:actin binding; IDA.  
DR GO:0005515; F:protein binding; IDA.  
DR GO:0016601; P:RAC protein signal transduction; IDA.  
DR InterPro; IPR002108; Actbind\_coflin.  
DR InterPro; IPR001452; SH3.  
DR Pfam; PF00241; cofilin\_ADF; 1.  
DR Pfam; PF00018; SH3; 1.  
DR PRINTS; PR00452; SH3DOMAIN.  
DR ProDom; PD000066; SH3; 1.  
DR SMART; SM00102; ADF; 1.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PSS0002; SH3; 1.  
SQ SEQUENCE 432 AA; 48341 MW; 4A36D3B6E59C707F CRC64;

Query Match 15.7%; Score 627.5; DB 11; Length 432;  
Best Local Similarity 58.0%; Pred. No. 9,7e-54;  
Matches 130; Conservative 13; Mismatches 44; Indels 37; Gaps 5;

QY 335 RARAPRAGTGVRCXVPVLNGARPERRPXKRSRTARSGT----- 377  
DB 229 RSAGAPRTG-----EPQEAIV--SRTQEWESAGQAHPREIFKQKER 270  
QY 378 WXTXSSISGPOPKLRSPFLQKLTQPTHTFGREPAAXSRPADLPABEPAPSPCLV 437  
DB 271 AMSTSVSSQPKLRSPFLQKLTQPTHTFGREPAAXSRPADLPABEPAPSPCLV-A 328  
QY 438 QABEAVYEEPEXQSTFYEQPPLVQOXXGSEHIDHITQGGLSXQGLCARALDYQAD 497  
DB 329 QREBPTEVPEODTLVEEPLVQOXXGSEHIDHITQGGLSXQGLCARALDYQAD 388  
QY 498 DXEISPDENLITGEVXXEGWNRGYPDGHFGMKPANYVELIE 541  
DB 389 TGEIVDEGWRGYPDGHFGMKPANYVELIE 432

## RESULT 9

Q9UM74 PRELIMINARY; PRT; 429 AA.  
AC Q9UM74; 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE SH3P74 (Fragment).  
GN SH3P74.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_Taxid=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yamazaki H., Shirao T.;  
RT "Rattus norvegicus mRNA for SH3P7 (SH3P7r4).";  
RC Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
DR EMBL; AB038364; BAA90866.1; -.  
DR HSSP; P19174; 2HSP.  
DR GO:0005622; C:intracellular; IEA.  
DR GO:0003779; F:actin binding; IEA.  
DR InterPro; IPR002108; Actbind\_coflin.  
DR InterPro; IPR001452; SH3.  
DR Pfam; PF00241; cofilin\_ADF; 1.  
DR Pfam; PF00018; SH3; 1.  
DR PRINTS; PR00452; SH3DOMAIN.  
DR ProDom; PD000066; SH3; 1.  
DR SMART; SM00102; ADF; 1.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PSS0002; SH3; 1.  
KW SH3 domain.  
SQ SEQUENCE 429 AA; 47923 MW; A7E7E7BFAF4A1705 CRC64;  
FT NON TER 1  
SQ SEQUENCE 429 AA; 47923 MW; A7E7E7BFAF4A1705 CRC64;

Query Match 15.2%; Score 606; DB 11; Length 429;

Best Local Similarity 59.9%; Pred. No. 1.3e-51;  
Matches 127; Conservative 13; Mismatches 58; Indels 14; Gaps 5;

QY 335 RARAPRAG-----TGVRCXVPVLNGARPERRPXKRSRTARSGTWRXTXSSISGPOP 389  
DB 227 RSAGAPRTGELGEVRSRQWESAGQA-PHREIFKQKERAMS-----TTSVSSQP 279  
QY 390 GKLRSPFLQKLTQPTHTFGREPAAXSRPADLPABEPAPSPCLVQABEAVYEEPX 449  
DB 280 GKLRSPFLQKLTQPTHTFGREPAAXSRPADLPABEPAPSPCLVQABEAVYEEPX 337  
QY 450 EDETFYEQPPLVQOXXGSEHIDHITQGGLSXQGLCARALDYQADDXEISPDENLI 509  
DB 338 EDETFYEQPPLVQOXXGSEHIDHITQGGLSXQGLCARALDYQADDXEISPDENLI 397  
QY 510 TGEVXXEGWNRGYPDGHFGMKPANYVELIE 541  
DB 398 TGEIVDEGWRGYPDGHFGMKPANYVELIE 429

## RESULT 10

Q9UM66 PRELIMINARY; PRT; 432 AA.  
AC Q9UM66; 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE SH3P73.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_Taxid=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yamazaki H., Shirao T.;  
RT "Rattus norvegicus mRNA for SH3P7 (SH3P7r3).";  
RC Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
DR EMBL; AB039819; BAA92709.1; -.  
DR HSSP; P19174; 2HSP.  
DR GO:0005622; C:intracellular; IEA.  
DR GO:0003779; F:actin binding; IEA.  
DR InterPro; IPR002108; Actbind\_coflin.  
DR InterPro; IPR001452; SH3.  
DR Pfam; PF00241; cofilin\_ADF; 1.  
DR Pfam; PF00018; SH3; 1.  
DR PRINTS; PR00452; SH3DOMAIN.  
DR ProDom; PD000066; SH3; 1.  
DR SMART; SM00102; ADF; 1.  
DR SMART; SM00326; SH3; 1.  
DR PROSITE; PSS0002; SH3; 1.  
KW SH3 domain.  
SQ SEQUENCE 432 AA; 48254 MW; E2D5E5C1D325B641 CRC64;

Query Match 15.1%; Score 603.5; DB 11; Length 432;

Best Local Similarity 59.0%; Pred. No. 2.4e-51;  
Matches 125; Conservative 13; Mismatches 61; Indels 13; Gaps 4;

QY 335 RARAPRAG-----TGVRCXVPVLNGARPERRPXKRSRTARSGTWRXTXSSISGPOP 389  
DB 229 RSAGAPRTGELGEVRSRQWESAGQA-PHREIFKQKERAMS-----TTSVSSQP 282  
QY 390 GKLRSPFLQKLTQPTHTFGREPAAXSRPADLPABEPAPSPCLVQABEAVYEEPX 449  
DB 283 GKLRSPFLQKLTQPTHTFGREPAAXSRPADLPABEPAPSPCLVQABEAVYEEPX 340  
QY 450 EDETFYEQPPLVQOXXGSEHIDHITQGGLSXQGLCARALDYQADDXEISPDENLI 509  
DB 341 EDETFYEQPPLVQOXXGSEHIDHITQGGLSXQGLCARALDYQADDXEISPDENLI 400  
QY 510 TGEVXXEGWNRGYPDGHFGMKPANYVELIE 541  
DB 401 TGEIVDEGWRGYPDGHFGMKPANYVELIE 432

```

RESULT 11
O9JMG7 PRELIMINARY; PRT; 433 AA.
AC O9JMG7;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE SH3P7r2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON NCBI_TaxID=10116;
RN 11
RP SEQUENCE FROM N.A.
RA Yamazaki H., Shitao T.;
RT "Rattus rattus mRNA for SH3P7 (SH3P7r2).";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AB039818; BA92708.1; -.
DR HSSP; P19174; ZHSP.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003779; F:actin binding; IEA.
DR InterPro; IPR002108; Actbind_cofln.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00241; coflin_ADF; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRODOM; PD000066; SH3; 1.
DR SMART; SM00102; ADF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 433 AA; 48341 MW; 952PFFB1E6A6EDDA CRC64;

Query Match 15.1%; Score 602; DB 11; Length 433;
Best Local Similarity 58.9%; Pred. No. 3,4e-51;
Matches 123; Conservative 12; Mismatches 44; Indels 30; Gaps 4;

QY 361 ERPPXKPSRT-----RSGWTNR-----XTXSISSPQPKL 392
DB 227 QHSAGPSPKRGELGEVVSRSQEWESAGQAPHPREIFKQERAMSTTSVSSQPKL 286
QY 393 RSPFLQKQLOPETHFGREPAAKXSRPADLPAREPAPSPCLVQAEPAVYEEFXEQE 452
DB 287 RSPFLQKQLOPETHFGREPAAKXSRPADLPAREPAPSPCLVQAEPAVYEEFXEQE 344
QY 453 TFYQPPPLVQOQXGSEHIDHHTQGGLSXQGLCARALYDYOAAADDEISFPDENLITGI 512
DB 345 TLYEPPPPVQPGAGSGHIDYMQSDLSGGGLCARALYDYOAAADDEISFPDENLITGI 404
QY 513 EVXLEGWVRGYGPDGHFGKMPANYVELIE 541
DB 405 EVIDEGWVRGYGPDGHFGKMPANYVELIE 433

RESULT 12
O9JHL4 PRELIMINARY; PRT; 436 AA.
AC O9JHL4;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE SH3P7r1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON NCBI_TaxID=10116;
RN 11
RP SEQUENCE FROM N.A.
RA Yamazaki H., Shitao T.;

```

```

RT "Rattus norvegicus mRNA for SH3P7 (SH3P7r1).";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN 12
RP SEQUENCE FROM N.A.
RA Yamazaki H., Shitao T.;
RT "SH3P7.";
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.
DR EMBL; AB038365; BA90867.1; -.
DR EMBL; AB009346; BA90819.1; -.
DR HSSP; P19174; ZHSP.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003779; F:actin binding; IEA.
DR InterPro; IPR002108; Actbind_cofln.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00241; coflin_ADF; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRODOM; PD000066; SH3; 1.
DR SMART; SM00102; ADF; 1.
DR SMART; SM00326; SH3; 1.
DR PROSITE; PS50002; SH3; 1.
KW SH3 domain.
SQ SEQUENCE 436 AA; 48612 MW; 064CA1E6AF84444 CRC64;

Query Match 15.0%; Score 599.5; DB 11; Length 436;
Best Local Similarity 56.4%; Pred. No. 6e-51;
Matches 127; Conservative 12; Mismatches 47; Indels 39; Gaps 5;

QY 354 LGNGARPERR-----PXKRPSPRT-----ARSGWTNR----- 379
DB 214 LQBAARQRQYQEGHRSAGPSPSRKGLGEVVSRSQEWESAGQAPHPREIFKQXE 273
QY 380 ---XTXSISSPQPKLRSPPLOKQLOPETHFGREPAAKXSRPADLPAREPAPSPCL 436
DB 274 RASTTTSVSSSQPKLSPFLQKQLOPETHFGREPAAKXSRPADLPAREPAPSPCL 331
QY 437 VQAEPAVYEEFXEQEFTFEPPPLVQOQXGSEHIDHHTQGGLSXQGLCARALYDYOAA 496
DB 332 AQTDEPFLYVPSQETLYEPPVQPGAGSGHIDYMQSDLSGGGLCARALYDYOAA 391
QY 497 DDXEISFPDENLITGIEVXKGMVRGYGPDGHFGKMPANYVELIE 541
DB 392 DDTETISFPDENLITGIEVIDEGWVRGYGPDGHFGKMPANYVELIE 436

RESULT 13
O7ZXQ9 PRELIMINARY; PRT; 447 AA.
AC O7ZXQ9;
DT 01-JUN-2003 (Tremblrel. 24, Created)
DT 01-JUN-2003 (Tremblrel. 24, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Similar to drebrin-like.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
OC Xenopodinae; Xenopus.
ON NCBI_TaxID=8355;
RN 11
RP SEQUENCE FROM N.A.
RA Kleins S., Strausberg R.;
RT Tissue=Embryo.
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC044296; AAH44296.1; -.
DR GO; GO:0005622; C:intracellular; IEA.
DR GO; GO:0003779; F:actin binding; IEA.
DR InterPro; IPR002108; Actbind_cofln.
DR InterPro; IPR001452; SH3.
DR Pfam; PF00241; coflin_ADF; 1.
DR Pfam; PF00018; SH3; 1.
DR PRINTS; PR00452; SH3DOMAIN.
DR PRODOM; PD000066; SH3; 1.

```

DR SMART; SM00102; ADF; 1.  
 DR SMART; SM00326; SH3; 1.  
 DR PROSITE; PSS0002; SH3; 1.  
 SQ SEQUENCE 447 AA; 50611 MW; 38C6767EE6E3B521 CRC64;

Query Match 9.2%; Score 365.5; DB 13; Length 447;  
 Best Local Similarity 45.8%; Pred. No. 1,2e-27;  
 Matches 81; Conservative 19; Mismatches 54; Indels 23; Gaps 6;

DB 386 SPOGKLRSPFLQXQLTQPTET-----HFGREPAAXSRERADLPABEAP--SXP-- 434  
 273 SAQGRRLRSPFLQXQACQSPSPSPRHRYGQEPSPVYPAHQTPESPVPVSHPPS 332

QY 435 -----CLV-QABEAVYEEPEXQETFEYEPPLVQ--QXGSEHIDHIOGGLSXOG 484  
 DB 333 TVHWKQCTASQOEENITQDATEDQNTYEDTTNNTYEDTPQEEPYEIEVE--EKG 390

QY 485 LCARALYDYOADADDEISFDPENITIGIEVXXEGMGVGPDGHFGKMPANYVELIE 541  
 DB 391 VCARALYDYOADADDEISFDPDLITQIFIDEGMKGFSAGHFGMPANYVELIE 447

RESULT 14  
 Q9VDF4 PRELIMINARY; PRT; 559 AA.

AC Q9VDF4; 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Contractin protein (LD29964P).  
 GN CORPACTIN OR CG3637.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NC NCB1\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams N.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amaratunga P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Barton R.C., Rogers J.H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Abayaratne A., An H.-J., Andrews-Pfannkoch L., Bessey E.M.,  
 RA Bailew R.M., Baer A., Baxendale J., Beldarrini D., Bolshakov S.,  
 RA Beeson K.Y., Bence P.V., Berman B.P., Bhattacharya P., Brothier P.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Burris K.A., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferreira C., Ferriera S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glaeser K.,  
 RA Glodex A., Gong F., Gottell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Heston D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalish P., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laake P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Matei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Mout S.M., Moy M., Murphy B., Mody P., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Ramirez K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wastaman D.A., Weinstein G.M., Weissbach U.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,

RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195 (2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Berkley;  
 RA Stapleton M., Brokstein P., Hong L., Abayaratne A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Fartan D., Frise E., George R.,  
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nuncio J., Pauley J., Paragas V., Park S., Pounanavong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Ceiniker S.,  
 RL Submitted (Aug-2001) to the EMBL/GenBank/DBJ databases.  
 CC -1- SIMILARITY: CONTAINS 1 SH3 DOMAIN.  
 DR EMBL; A8003733; AAF55840.1;  
 DR EMBL; AY051774; AAK93198.1;  
 DR HSSP; P29355; ISEM.  
 DR FlyBase; FBgn025865; Contractin.  
 DR InterPro; IPR003134; Hs1/Contractin.  
 DR InterPro; IPR00108; Neu\_cyt\_fact\_2.  
 DR InterPro; IPR001452; SH3.  
 DR Pfam; PF02218; Hs1\_rep; 4.  
 DR Pfam; PF00018; SH3; 1.  
 DR PRINTS; PR00499; P67PHOX.  
 DR PRINTS; PR00452; SH3DOMAIN.  
 DR ProDom; PD000066; SH3; 1.  
 DR SMART; SM00326; SH3; 1.  
 DR PROSITE; PSS0002; SH3; 1.  
 KM SH3 domain.  
 SQ SEQUENCE 559 AA; 61083 MW; 2B767BD3EFF1BCB2 CRC64;

Query Match 5.8%; Score 231; DB 5; Length 559;  
 Best Local Similarity 27.1%; Pred. No. 4e-14;  
 Matches 77; Conservative 43; Mismatches 110; Indels 54; Gaps 12;

QY 291 SKESRP--GTRRQGRGASVGLGXPKXRL-----SPAGRPPLHAPRG----- 334  
 DB 292 TSTAPPKSGSRRAITGRGGIGNALSAPNMQSVSTPPARKPILITPKAPKIKL 351

QY 335 RARAPPAGTGVACXVPVLGNAR--PERRPXXRPSRTASWGTWXTYSSISPOGKLR 393  
 DB 352 EAKEEPASTSAVAAPTPVPAARPEPAVAKAAPDPDVPQIEVETVTPRPSRQ 411

QY 394 SPTLQXQLTQPTETFGREPAAXSRERADLPABEAPSPKPCUYOABEAVYEEPEXQET 453  
 DB 412 SP-VVYPTPQPEVH-----AQVQVQ-----EPQPADPEPV-VVEEPLYONQAEIKR 457

QY 454 FYEQP-----LVQOQX--XGSEHIDHIOGGLSXOGLCARALYDYOAA 496  
 DB 458 ASPLPFPNGVSAVAAPSGTATVPEEAYVANSNDLADYLEDTSIH-----ALALYDYOAA 512

QY 497 DDXEISFDPENITIGIEVXXEGMGVGPDGHFGKMPANYVELI 540  
 DB 513 DDEISFPDPDVTIHEKIDGWRGLCKN-RYGLPFPANYVQV 555

RESULT 15  
 Q92218 PRELIMINARY; PRT; 486 AA.

AC Q92218; 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
 DE Hematopoietic cell specific lym substrate 1.  
 GN HCSL1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathu; Muridae; Murinae; Mus.  
 NC NCB1\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAUBERG R.;



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2004, 10:23:34 ; Search time 53.1947 Seconds

(without alignments)  
4164.275 Million cell updates/sec

Title: US-10-028-952a-10

Perfect score: 3752  
Sequence: 1 IRHEHGEERTYKXNKXAPG.....DXAIALQPKXERTLSQKX 784

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: Genesep29Jan04:\*  
2: Genesep2980s:\*  
3: Genesep2990s:\*  
4: Genesep2000s:\*  
5: Genesep2001s:\*  
6: Genesep2002s:\*  
7: Genesep2003s:\*  
8: Genesep2004s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query %	Length	DB ID	Description
1	3251.5	86.7	753	4	AA666395 Human mam
2	1177.5	31.4	760	4	AA666394 Human pro
3	435	11.6	538	2	AA123756 Human mam
4	238	6.3	55	3	AA653794 Human col
5	199	5.2	40	3	AA653737 Human col
6	194.5	5.2	91	3	AA653743 Human col
7	185	4.9	72	3	AA653817 Human col
8	149	4.0	7052	6	AA653826 Human P45
9	145	3.9	54	4	AA653774 Human col
10	145	3.9	54	4	AA653774 Human col
11	138	3.7	50	3	AA653773 Human col
12	122	3.3	123	7	AA653956 Human bre
13	122	3.3	123	7	AA653957 Human bre
14	120	3.2	96	4	AA653827 Human col
15	120	3.2	96	4	AA653827 Human col
16	120	3.2	96	4	AA653827 Human col
17	120	3.2	96	4	AA653827 Human col
18	120	3.2	96	4	AA653827 Human col
19	119	3.2	97	5	AA653827 Human col
20	119	3.2	97	5	AA653827 Human col
21	119	3.2	97	5	AA653827 Human col
22	119	3.2	97	5	AA653827 Human col
23	117	3.1	106	5	AA653827 Human col
24	116	3.1	139	3	AA653822 Human col
25	115	3.1	47	3	AA653894 Human col

26	113.5	3.0	19938	6	AA688398 Streptomy
27	113	3.0	82	4	AA685047 Human imm
28	113	3.0	386	2	AA685047 Human imm
29	111	3.0	673	6	AA684009 Human POM
30	110.5	2.9	80	4	AA684574 Human imm
31	110	2.9	465	2	AA682653 Human cal
32	110	2.9	2087	4	AA683156 Human aci
33	109.5	2.9	19938	6	AA682681 Streptomy
34	109.5	2.9	19938	6	AA682681 Streptomy
35	108	2.9	2091	3	AA682000 Human pro
36	106.5	2.8	476	4	AA6803460 Novel hum
37	106.5	2.8	476	4	AA6803460 Novel hum
38	106	2.8	51	6	AA6803864 Human mus
39	106	2.8	51	6	AA6803864 Human mus
40	105.5	2.8	70	4	AA682686 Human dig
41	105	2.8	221	4	AA683924 Human pro
42	105	2.8	19938	6	AA682681 Streptomy
43	104.5	2.8	1212	2	AA687503 Human N-m
44	104	2.8	42	4	AA687503 Human N-m
45	103.5	2.8	51	4	AA686866 Human imm

## ALIGNMENTS

RESULT 1  
ID AA666395 standard; protein; 753 AA.

10-APR-2001 (first entry)  
Human mammatstatin C protein sequence.

Human mammatstatin A; mammatstatin B; mammatstatin C; ECGI; cancer;  
epithelial cell growth inhibitor.

Homo sapiens.

WO200078955-A1.

19-JUN-2000; 2000WO-US016900.

18-JUN-1999; 99US-0139995P.

(BIOT-) BIOTHERAPIES INC.

WPI; 2001-071393/08.

A Mammatstatin-like epithelial cell growth inhibitor, and the nucleic that  
encodes it, useful for diagnosing and/or preventing epithelial cell  
cancers, e.g. of the ovaries or prostate.

Disclosure; Page 44-46; 55pp; English.

The present invention describes a mammatstatin-like epithelial cell growth  
inhibitor (ECGI) which has substantial similarity to mammatstatin A, B or  
C. ECGI is expressed in healthy cells, but is either absent or expressed  
at reduced levels in cancerous cells. The protein and its coding sequence  
can be used to inhibit epithelial cell growth and the amount present in  
cells can be used to diagnose cancer or monitor its progress

Sequence 753 AA:

Query Match 86.7%; Score 3251.5; DB 4; Length 753;  
Best Local Similarity 96.0%; Pred. No. 0;  
Matches 753; Conservative 0; Mismatches 0; Indels 31; Gaps 31;



```

QY 1 IRHEGEETVEYNKKEAFGAPVPVSPRGARCGRRPCPPVYKXYSXDXTPVRRGGE 60
DB 1 IRHEGEETVEYNK- EAPGAPVPVSPRGARCGRRPCGPVYKXYSR- XTDPVRRGGE 58
QY 61 RGLAAGARPARPGATRGSGASAWGYLGRYTCQTCQYQVSKCELEDRRLPMKRRPAK 120
DB 59 RGLA- GAKRPARRRGATRGSGASAWGYLGRYTCQ- XQVSKCELEDRRLPM- RRAK 115
QY 121 ARLILFSTNTDESASRSFDPFGYAXXVRRKTTGTGLMRPSXHSVAAVXSDVSS 180
DB 116 ARLILFSTNTDESASRSFDPFGYAX- XVRKTTGTGLMRPS- HSDVAXSDVSS 173
QY 181 YHAEFTKRWIHPILIGXSSWYTVVROVSFTLLMCCCHGNPAQYENRERXRLVYL 240
DB 174 YHAEFTKRWIHPILIG- XSWYTVVROVSFTLL- MCCCHGNPAQYENRERXRLVYL 231
QY 241 GXGANGAKXXSVGLXINASKSESPYGTIRORRGASVGLGXDPRLSPAGRPPESTRX 300
DB 232 GXGANGAK- XSVGLXINASKSESRP- GTIRORRGASVGLGXDPRLSPAGRPPESTR- X 288
QY 301 RAGGRVPRAPRGSGXAECPSSWETGRGKGPILARHAPYRARAEFXSSSTINRHTS 360
DB 289 RAGGRVPRAPRGSGS- AECPSWETGRGKGPILARHAPYRARAEF- XSTINRHTS 346
QY 361 ACIFVXXXIILFLWVDIQXWDCXXTWXFFWFIKESYXXXSLYKETSLLPSDPFKXERYV 420
DB 347 ACIFV- XXILFLWVDIQXWDCXXTWXFFWFIKESY- XXXSLYKETSLLPSDPFK- ERYV 403
QY 421 WRAPHXYPFTYLLQNDPFGYRYLOVSKXROIEYXNFCIRGDFYIOSCMKDKCSRDL 480
DB 404 WRAPHXYPFTYLLQNDPFGYRYLOVS- XROIEYXNFCIRGDF- IOSCMKDKCSRDL 461
QY 481 QSNWMSOMXYISSSTSSXSTELALXSLIPYXXXXXGFI SNILXGIKIXXVYLFSL 540
DB 462 QSNWMSOMXYISSSTSS- XSTELALXSLIPY- XXXXGFI SNILXGIKIXXVYLFSL 519
QY 541 AAFYONIKXXPSIGHYCTRHCVCHSKFPMWCSQXFCRVXSLTVVPLFGRNLTXX 600
DB 520 AAFYONIK- XPSIGHYCTRHCVCH- SKFPMWCSQXFCRVXSLTVVPLFGRNLT- X 576
QY 601 FNRKXSNRRTYXXITLQISPHYASTCACYLIGSCYFPYFLSLXTPFPSPHPS 660
DB 577 FNRKXSNRRTYX- ITLQISPHYASTCACYLIGSCYFPYFLSL- XTPFPSPHPS 634
QY 661 FPIIVYITNTCLSEQLIHKRXSSTGEEXLIPVILALXAKGRLSBRVNDQEPQGE 720
DB 635 FPIIV- YITNTCLSEQLIHKRXSSTGEEXLIPVILAL- XAKGRLSBRVNDQEP- QGE 691
QY 721 TLSIQKTKIXPYVLAHTCSLSYSEGXXIXDAOEVEAAVXDXAIALQFEXERETLS 780
DB 692 TLSIQKTKIXPYVLAHTCSLSYSEGXX- XIDAEVEAAVXDX- AIALQFEXERETLS 749
QY 781 QKXK 784
DB 750 QKXK 753

```

RESULT 2  
AAB6394  
ID AAB6394 standard; protein; 760 AA.

AC AAB6394;  
XX  
DT 10-APR-2001 (first entry)  
XX Human prostate ECGI protein sequence.  
XX Human; mammatstatin A; mammatstatin B; mammatstatin C; ECGI; cancer;  
XX epithelial cell growth inhibitor.  
XX Homo sapiens.  
XX  
PN WO200078955-A1.

```

XX 28-DEC-2000.
PD
XX 19-JUN-2000; 2000MO-US016900.
XX
PF 18-JUN-1999; 99US-0139995P.
XX
PR (BIOT-) BIOTHERAPIES INC.
XX
PA
XX
PI Eryn PR;
XX
XX WPI; 2001-071393/08.
XX
XX N-PSDB; AAF31282.
XX
XX A Mammatstatin-like epithelial cell growth inhibitor, and the nucleic that
XX PT encodes it, useful for diagnosing and/or preventing epithelial cell
XX PT cancers, e.g. of the ovaries or prostate.
XX
XX Disclosure; Page 41-43; 55pp; English.
XX
XX The present invention describes a mammatstatin-like epithelial cell growth
XX CC inhibitor (ECGI) which has substantial similarity to mammatstatin A, B or
XX CC C. ECGI is expressed in healthy cells, but is either absent or expressed
XX CC at reduced levels in cancerous cells. The protein and its coding sequence
XX CC can be used to inhibit epithelial cell growth and the amount present in
XX CC cells can be used to diagnose cancer or monitor its progress
XX
XX Sequence 760 AA;

```

Query Match 31.4%; Score 1177.5; DB 4; Length 760;  
Best local similarity 81.7%; Pred. No. 1.9e-120;  
Matches 267; Conservative 4; Mismatches 39; Indels 17; Gaps 14;

```

QY 5 HGEETVEYNKKEAFGAPVPVSPRGARCGRRPCGPVYKXYSXDXTPVRRGGEPRGAL 64
DB 37 HGEETVEYNKKEAFGAPVPVSPRGARCG- XRPCGPVYKXYSR- FTDVRRGGEPRGAL 94
QY 65 AXGAKRPARPGATRGSGASAWGYLGRYTCQTCQYQVSKCELEDRRLPMKRRPAK 124
DB 95 ASXAKRPARPGATRG- XARWGYLGRYTCQ- QVSKCELEDRRLPMKRRPAK 152
QY 125 LIFSTNTDESASRSFDPFGYAXXVRRKTTGTGLMRPSXHSVAAVXSDVSSYTHX 164
DB 153 LIFSTNTD- XSGASRSFDPFGYAX- GVRKTTGTGLMRPSVSDVAFSFDVSSYTH- X 209
QY 185 AEFTKRWIHPILIGXSSWYTVVROVSFTLLMCCCHGNPAQYENRERXRLVYL 244
DB 210 AEFTKRWIHPILIG- NSHDTTVVROVSFTLLMCCCHGNPAQYENRER- XHLVYVVLGXGA 267
QY 245 NGAXKXSVGLXINASKSESPYGTIRORRGASVGLGXDPRLSPAGRPPESTRX 303
DB 268 NGAK- LSVGLXINASKSESPYGTIRORRGASVGLGX- RLSPAGRPPESTRX 324
QY 304 GRVPRARPGSGXAECPSSWETGRGGRK 330
DB 325 RAEPRA--GRVRCXVYFVLGANGAR 347

```

RESULT 3  
AA23756  
ID AA23756 standard; protein; 538 AA.

AC AA23756;  
XX  
DT 09-SEP-1999 (first entry)  
XX Human mammatstatin amino acid sequence.  
XX Human; mammatstatin; breast cancer.  
XX Homo sapiens.  
XX  
PN WO9932625-A2.

XX 01-JUL-1999.  
 XX 18-DEC-1998; 98MO-US027147.  
 XX 19-DEC-1997; 97US-00994076.  
 XX (UNMI ) UNITV MICHIGAN.  
 XX Ervin PR;  
 XX WPI: 1999-418924/35.  
 XX N-PSDB; AAB55833.  
 XX DNA encoding Mammatstatin, a specific mammary cell growth inhibitor,  
 XX useful for therapy of breast cancer.  
 XX Claim 2; Page 45-47; 77pp; English.  
 XX The present sequence represents human mammatstatin. Mammatstatin produced  
 XX by normal human mammary cells in culture and recombinant expressed  
 XX CC Mammatstatin inhibit the growth of human mammary carcinoma cells, and is  
 XX CC useful as a therapeutic agent in the treatment of breast cancer  
 XX  
 XX Sequence 538 AA;

Query Match 11.6%; Score 435; DB 2; Length 538;  
 Best Local Similarity 28.8%; Pred. No. 1,1e-38;  
 Matches 182; Conservative 50; Mismatches 216; Indels 184; Gaps 30;

QY 123 LILFSTINTDESASRSPFGFYAXXVRYTGTITGLMPSPXHSDDVAFKSPDVGSSYH 182  
 DB 1 MILFSTINTDESASGSS-----DLTFWLSRRCC-----KSYH 33  
 QY 183 XEAFTKXWIV-----HPLIGXXSWXTVV-----EQVS----- 211  
 DB 34 RD-----NWLVAAKRSKRRLILRCRLFLSLGSIHQALDCSPTNRERELGLDRRTGL 88  
 QY 212 FTLLMCCCHGNPAQYERNRRLVYVLGXGANGAKXXSVGLXLNKSKSRPXTIRQ 271  
 DB 89 FTLLMCCCHGNPAQYERNRRLVYVLGXGANGAKLPVGLLNKSKSRPXTIRQ 148  
 QY 272 RRGASVGLGXPRPRLSPAGRPBPSTRXKXAGRVPRPAGPAGSABCPSSWETGRGRK 331  
 DB 149 RRGASVGLGXPRPRLSPAGRPBPSTRXKXAGRVPRPAGPAGSABCPSSWETGRGRK 181  
 QY 332 GXPLARH-----APHVABAEFXXSSTIHNHTSACIFMXXXILFLVNDIQWDCX 382  
 DB 182 ACPAARRRRGVRSALRFGKGAAGKAAPSPVTHRTFVLVNSARVAFITIDIV-HVS 240  
 QY 383 XTWYFVWF-----IEKSSYXXXXRLYKFTSPSPDFCKERMWENAPHPXP-PTXL 434  
 DB 241 LLVNEFFSFGIESSGIGARSPGSSISGLLR--NLHTDFH-----RGCTMLHPYQVIFL 291  
 QY 435 LQNFQKRYVQVEXXRGIEYXNFCIRGT---DPIQSCMKKDCSSDDLQSYMKSGMKY 491  
 DB 292 NMKMSGEMPL-ISTPLPLPYCRMTSRGTGYKPHYTDKNITETISALEAQILGK----- 345  
 QY 492 ISSSTTSXXST-----ELALXSLI--PTY-----XXXXXGPFISNITLXG 528  
 DB 346 VVGRITSLALGTCKAGIGNLRLANTFLVVPAYILLNMLCDHLLYLITKNEGLVQIYFKV 405  
 QY 529 IKIKKXVVLFSIAPFQNIKXXPSIGLHY-----CTHCCHXSKXKPSWQSCQ 576  
 DB 406 KKSNNRKSCTLLHNFETKXNSRLLG--ITYVPDFTVFTFOKCS--HGVANNV--GW--- 458  
 QY 577 XFCRVKSLTVVRLFSKKNLTXKXENLRKVSNRTRTYXLTITQISPYHTAS--TCACXL 634  
 DB 459 -----RNSLTIVRLF--RNETSEPM-----SSTCAKLVGTGLPEPESHRSLSHTLLAH 506  
 QY 635 IPGSGYF-PFYFLSLXITTPSPHPFSPFLIV 665  
 DB 507 VPVILFLAPKLFPPSSSHNPPFPFL-FFLFI 537

RESULT 4  
 AAB53794  
 ID AAB53794 standard; protein; 55 AA.  
 XX AC  
 XX AAB53794;  
 XX DT  
 XX 09-MAR-2001 (first entry)  
 XX DE  
 XX Human colon cancer antigen protein sequence SEQ ID NO:1334.  
 XX KW Human; colon cancer; colon cancer antigen; diagnosis; detection;  
 XX KW identification; cytostatic; cardioactive; neuroprotective; vulnerary;  
 XX KW immunomodulatory; muscular; gynaecological; gastrointestinal;  
 XX KW nephrotropic; antiinfective; antibacterial; gene therapy; wound;  
 XX KW neural disorder; immune system disorder; muscular disorder;  
 XX KW reproductive disorder; gastrointestinal disorder; renal disorder;  
 XX KW infectious disease; cardiovascular disorder.  
 XX OS  
 XX Homo sapiens.  
 XX PN  
 XX WC0200055351-A1.  
 XX PD  
 XX 21-SEP-2000.  
 XX PF  
 XX 08-MAR-2000; 2000MO-US005883.  
 XX PR  
 XX 12-MAR-1999; 99US-0124270P.  
 XX PA  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX PI  
 XX Rosen CA, Ruben SM;  
 XX DR  
 XX WPI, 2000-587534/55.  
 XX N-PSDB; AAC98551.  
 XX PT  
 XX Colon cancer associated gene sequences, referred to as colon cancer  
 XX PT antigens, useful for the treatment, prevention, and diagnosis of colon  
 XX PT disorders such as colon cancer.  
 XX PS  
 XX Claim 11; Page 1915; 2104pp; English.  
 XX CC  
 XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,  
 XX CC called human colon cancer antigens, given in AAB53234 to AAB54006. The  
 XX CC human colon cancer antigens can have cytostatic, cardioactive, muscular,  
 XX CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,  
 XX CC vulnerary, nephrotropic, antiinfective and antibacterial activities, and  
 XX CC can be used in gene therapy. The colon cancer antigen polynucleotides,  
 XX CC proteins and antibodies to the proteins are useful for the prevention,  
 XX CC treatment and diagnosis of colon disorders, such as colon cancer. The  
 XX CC polynucleotides may be used in diagnostics and research, such as for  
 XX CC chromosome identification, and as hybridisation probes. The proteins may  
 XX CC also be used to prevent diseases such as neural disorders, immune system  
 XX CC disorders, muscular disorders, reproductive disorders, gastrointestinal  
 XX CC disorders, wounds, renal disorders, infectious diseases, and  
 XX CC cardiovascular disorders. AAC98764 to AAC98772 and AAB54007 represent  
 XX CC sequences used in the exemplification of the present invention  
 XX  
 XX Sequence 55 AA;

Query Match 6.3%; Score 238; DB 3; Length 55;  
 Best Local Similarity 86.5%; Pred. No. 3.5e-18;  
 Matches 45; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 37 CGPPVYKHYSDXXTDPVRRGEPGRGALAXGAKRPAARPGATRSQXGARWG 88  
 DB 2 CGPPVYKHYSDXXTDPVRRGEPGRGALAXGAKRPAARPGATRSQXGARWG 53

RESULT 5  
 AAB53737  
 ID AAB53737 standard; protein; 40 AA.

XX AAB53737;  
 AC 09-MAR-2001 (first entry)  
 DT Human colon cancer antigen protein sequence SEQ ID NO:1277.  
 DE Human; colon cancer; colon cancer antigen; diagnosis; detection;  
 XX identification; cytostatic; cardiostatic; neuroprotective; vulnary;  
 XX immunomodulatory; muscular; gynaecological; gastrointestinal;  
 XX nephrotoxic; antineoplastic; antibacterial; gene therapy; wound;  
 XX neutral disorder; immune system disorder; muscular disorder;  
 XX reproductive disorder; gastrointestinal disorder; renal disorder;  
 XX infectious disease; cardiovascular disorder.  
 XX Homo sapiens.  
 OS WO200055351-A1.  
 PN 21-SEP-2000.  
 XX 08-MAR-2000; 2000WO-US005883.  
 XX 12-MAR-1999; 99US-0124270P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Ruben SM;  
 XX WPI: 2000-587534/55.  
 XX N-PSDB; AAC98494.  
 DR Colton cancer associated gene sequences, referred to as colon cancer  
 PT antigens, useful for the treatment, prevention, and diagnosis of colon  
 PT disorders such as colon cancer.  
 XX Claim 11; Page 1864; 2104pp; English.  
 XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,  
 CC called human colon cancer antigens, given in AAB53234 to AAB54006. The  
 CC human colon cancer antigens can have cytostatic, cardiostatic, muscular,  
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,  
 CC vulnary, nephrotoxic, antineoplastic, antibacterial activities, and  
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,  
 CC proteins and antibodies to the proteins are useful for the prevention,  
 CC treatment and diagnosis of colon disorders, such as colon cancer. The  
 CC polynucleotides may be used in diagnostics and research, such as for  
 CC chromosome identification, and as hybridisation probes. The proteins may  
 CC also be used to prevent diseases such as neural disorders, immune system  
 CC disorders, muscular disorders, reproductive disorders, gastrointestinal  
 CC disorders, wounds, renal disorders, infectious diseases, and  
 CC cardiovascular disorders. AAC98764 to AAC98772 and AAB54007 represent  
 CC sequences used in the exemplification of the present invention  
 CC  
 XX SQ Sequence 40 AA;  
 QY Query Match 5.3%; Score 199; DB 3; Length 40;  
 DB Best Local Similarity 90.0%; Pred. No. 4.8e-14;  
 Matches 36; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 202 WVTWVGVSTFLXMCCHGNPAQYERNRXRLVTVUG 241  
 DB 1 WVTWVGVSTFLXMCCHGNPAQYERNRXRLVTVUG 40  
 RESULT 6  
 ID AAB53743 standard; protein; 91 AA.  
 XX AAB53743;  
 AC AAB53743;  
 XX 09-MAR-2001 (first entry)  
 DT Human colon cancer antigen protein sequence SEQ ID NO:1277.  
 XX

DE Human colon cancer antigen protein sequence SEQ ID NO:1283.  
 XX Human; colon cancer; colon cancer antigen; diagnosis; detection;  
 XX identification; cytostatic; cardiostatic; neuroprotective; vulnary;  
 XX immunomodulatory; muscular; gynaecological; gastrointestinal;  
 XX nephrotoxic; antineoplastic; antibacterial; gene therapy; wound;  
 XX neutral disorder; immune system disorder; muscular disorder;  
 XX reproductive disorder; gastrointestinal disorder; renal disorder;  
 XX infectious disease; cardiovascular disorder.  
 XX Homo sapiens.  
 OS WO200055351-A1.  
 PN 21-SEP-2000.  
 XX 08-MAR-2000; 2000WO-US005883.  
 XX 12-MAR-1999; 99US-0124270P.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 XX Rosen CA, Ruben SM;  
 XX WPI: 2000-587534/55.  
 XX N-PSDB; AAC98500.  
 DR Colton cancer associated gene sequences, referred to as colon cancer  
 PT antigens, useful for the treatment, prevention, and diagnosis of colon  
 PT disorders such as colon cancer.  
 XX Claim 11; Page 1868-1869; 2104pp; English.  
 XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,  
 CC called human colon cancer antigens, given in AAB53234 to AAB54006. The  
 CC human colon cancer antigens can have cytostatic, cardiostatic, muscular,  
 CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,  
 CC vulnary, nephrotoxic, antineoplastic, antibacterial activities, and  
 CC can be used in gene therapy. The colon cancer antigen polynucleotides,  
 CC proteins and antibodies to the proteins are useful for the prevention,  
 CC treatment and diagnosis of colon disorders, such as colon cancer. The  
 CC polynucleotides may be used in diagnostics and research, such as for  
 CC chromosome identification, and as hybridisation probes. The proteins may  
 CC also be used to prevent diseases such as neural disorders, immune system  
 CC disorders, muscular disorders, reproductive disorders, gastrointestinal  
 CC disorders, wounds, renal disorders, infectious diseases, and  
 CC cardiovascular disorders. AAC98764 to AAC98772 and AAB54007 represent  
 CC sequences used in the exemplification of the present invention  
 CC  
 XX SQ Sequence 91 AA;  
 QY Query Match 5.2%; Score 194.5; DB 3; Length 91;  
 DB Best Local Similarity 59.2%; Pred. No. 4.1e-13;  
 Matches 42; Conservative 3; Mismatches 15; Indels 11; Gaps 2;  
 QY 284 PRISPAPGRP-----PPSTRXXRAGGRVRRAPGPGSXAECSSWETGRGKXPL 335  
 DB 19 PRISPAPGRP-----PPSTRXXRAGGRVRRAPGPGSXAECSSWETGRGKXPL 75  
 QY 336 ARHAPHYRARA 346  
 DB 76 RLDPSIRTRA 86  
 RESULT 7  
 ID AAB53817 standard; protein; 73 AA.  
 XX AAB53817;  
 AC AAB53817;  
 XX 09-MAR-2001 (first entry)  
 DT Human colon cancer antigen protein sequence SEQ ID NO:1357.  
 XX

```

XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardioactive; neuroprotective; vulnery;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotoxic; anti-infective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder.
XX
XX Homo sapiens.
XX
XX WO200055351-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US005883.
XX
XX 12-MAR-1999; 99US-0124270P.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Rosen CA, Ruben SM;
XX
XX WPI: 2000-587534/55.
XX
XX N-PSDB; AAC98574.
XX
XX Colon cancer associated gene sequences, referred to as colon cancer
XX antigens, useful for the treatment, prevention, and diagnosis of colon
XX disorders such as colon cancer.
XX
XX Claim 11; Page 1932-1933; 2104pp; English.
XX
XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,
XX called human colon cancer antigens, given in AAB53234 to AAB54006. The
XX human colon cancer antigens can have cytostatic, cardioactive, muscular,
XX neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
XX vulnery, nephrotoxic, anti-infective and antibacterial activities, and
XX can be used in gene therapy. The colon cancer antigen polynucleotides,
XX proteins and antibodies for the proteins are useful for the prevention,
XX treatment and diagnosis of colon disorders, such as colon cancer. The
XX polynucleotides may be used in diagnostics and research, such as for
XX chromosome identification, and as hybridisation probes. The proteins may
XX also be used to prevent diseases such as neural disorders, immune system
XX disorders, muscular disorders, reproductive disorders, gastrointestinal
XX disorders, wounds, renal disorders, infectious diseases, and
XX cardiovascular disorders. AAC98764 to AAC98772 and AAB54007 represent
XX sequences used in the exemplification of the present invention
XX
XX
XX Sequence 73 AA;
SQ
Query Match 4.9%; Score 185; DB 3; Length 73;
Best Local Similarity 69.1%; Pred. No. 3.5e-12;
Matches 38; Conservative 2; Mismatches 7; Indels 8; Gaps 1;
QY 284 PLSPSPAGRP-----PSTRTXXRAGGRVRRRPPGSGXAECCSSMETGRGR 330
DB 19 PRLASPSRSPACPRRRAPXSTRXARAGGRVRRRPPGSGAECCSSMETGRGR 73

```

```

OS Homo sapiens.
XX
XX WO200295034-A2.
XX
XX 28-NOV-2002.
XX
XX 23-MAY-2002; 2002WO-CA000758.
XX
XX 23-MAY-2001; 2001US-0292531P.
XX
XX (CYTO-) CYTOCHROME INC.
XX
XX Wistniewski J, Pekovich PM, Ramshaw H;
XX
XX WPI: 2003-112152/10.
XX
XX N-PSDB; AAD53861.
XX
XX New isolated P450RAI-3 nucleic acid molecule and polypeptide, useful for
XX diagnosing, preventing, or treating disorders with aberrant expression or
XX activity of the P450RAI-3, such as cancer, actinic keratosis, acne and
XX psoriasis.
XX
XX Example 1; Fig 1; 231pp; English.
XX
XX The invention relates to retinoic acid metabolising cytochrome P450,
XX P450RAI polypeptides and polynucleotides. Methods and compositions of the
XX invention are useful for diagnosing, preventing, ameliorating and/or
XX treating disorders associated with the aberrant expression or activity of
XX the P450RAI such as diseases related to vitamin A and retinoic acid
XX metabolism, e.g., cancer, actinic keratosis, psoriasis, acne and
XX ichthyosis. The invention is useful in gene therapy and as vaccine. The
XX present sequence is human P450RAI protein
XX
XX
XX Sequence 7052 AA;
SQ
Query Match 4.0%; Score 149; DB 6; Length 7052;
Best Local Similarity 54.2%; Pred. No. 9.7e-06;
Matches 39; Conservative 4; Mismatches 25; Indels 4; Gaps 3;
QY 713 DOPXGGETSLQNTKIKXPVLAFHTCSLSYSEGGXGXIIDXAQEVAAAVRXDXAIALQP 772
DB 427 DOPDQGETSLKNTKIKXPVLAFHTCSLSYSEGGXGXIIDXAQEVAAAVRXDXAIALQP 482
QY 773 GXERETLSQKX 784
DB 483 GQOSETPSQKX 494

```

```

RESULT 9
AAB53774
ID AAB53774 standard; protein; 54 AA.
AC AAB53774;
XX
XX 09-MAR-2001 (first entry)
XX
XX Human colon cancer antigen protein sequence SEQ ID NO:1314.
XX
XX Human; colon cancer; colon cancer antigen; diagnosis; detection;
KW identification; cytostatic; cardioactive; neuroprotective; vulnery;
KW immunomodulatory; muscular; gynaecological; gastrointestinal;
KW nephrotoxic; anti-infective; antibacterial; gene therapy; wound;
KW neural disorder; immune system disorder; muscular disorder;
KW reproductive disorder; gastrointestinal disorder; renal disorder;
KW infectious disease; cardiovascular disorder.
XX
XX Homo sapiens.
XX
XX WO200055351-A1.
XX
XX 21-SEP-2000.
XX
XX 08-MAR-2000; 2000WO-US005883.
XX

```

XX 12-MAR-1999; 99US-0124270P.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Ruben SM;  
XX WPI; 2000-597534/55.  
XX N-PSDB; AAC98531.  
XX Colton cancer associated gene sequences, referred to as colon cancer  
PT antigens, useful for the treatment, prevention, and diagnosis of colon  
PT disorders such as colon cancer.  
BS Claim 11; Page 1896; 2104P; English.  
XX AAC97991 to AAC98763 encode the human colon cancer associated proteins,  
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The  
CC human colon cancer antigens can have cytostatic, cardiostatic, muscular,  
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,  
CC vulnerability, antiproliferative, anti-infective and antibacterial activities, and  
CC can be used in gene therapy. The colon cancer antigen polynucleotides,  
CC proteins and antibodies to the proteins are useful for the prevention,  
CC treatment and diagnosis of colon disorders, such as colon cancer. The  
CC polynucleotides may be used in diagnostics and research, such as for  
CC chromosome identification, and as hybridization probes. The proteins may  
CC also be used to prevent diseases such as neural disorders, immune system  
CC disorders, muscular disorders, reproductive disorders, gastrointestinal  
CC disorders, wounds, renal disorders, infectious diseases, and  
CC cardiovascular disorders. AAC98764 to AAC98772 and AAC98773 represent  
CC sequences used in the exemplification of the present invention  
XX Sequence 54 AA;  
SQ  
Query Match 3.9%; Score 145; DB 3; Length 54;  
Best Local Similarity 96.3%; Pred. No. 6.3e-08;  
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
Cy 205 TTVAGVSEFTLXMCCHGNPAQYERNR 231  
Db 1 TTVAGVSEFTLXMCCHGNPAQYERNR 27  
RESULT 10  
AAC96009  
ID AAM96009 standard; protein; 54 AA.  
XX AAM96009;  
AC AAM96009;  
XX 21-NOV-2001 (first entry)  
DT 21-NOV-2001 (first entry)  
XX Human reproductive system related antigen SEQ ID NO: 4667.  
DB Human reproductive system related antigen; reproductive system disorder;  
XX Human; reproductive system related antigen; reproductive system disorder;  
KW Cancer; gene therapy.  
XX Homo sapiens.  
OS Homo sapiens.  
XX WO200155320-A2.  
PN WO200155320-A2.  
XX 02-AUG-2001.  
PD 02-AUG-2001.  
XX 17-JAN-2001; 2001WO-US001339.  
PF 17-JAN-2001; 2001WO-US001339.  
XX 31-JAN-2000; 2000US-0179065P.  
PR 04-FEB-2000; 2000US-0180628P.  
PR 24-FEB-2000; 2000US-0184664P.  
PR 02-MAR-2000; 2000US-0186350P.  
PR 16-MAR-2000; 2000US-0189874P.  
PR 17-MAR-2000; 2000US-0190076P.  
PR 18-APR-2000; 2000US-0198123P.  
PR 19-MAY-2000; 2000US-0205515P.  
PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.  
PR 30-JUN-2000; 2000US-0215135P.  
PR 07-JUL-2000; 2000US-0216647P.  
PR 07-JUL-2000; 2000US-0216680P.  
PR 11-JUL-2000; 2000US-0217487P.  
PR 11-JUL-2000; 2000US-0217496P.  
PR 14-JUL-2000; 2000US-0218390P.  
PR 26-JUL-2000; 2000US-0220963P.  
PR 26-JUL-2000; 2000US-0220964P.  
PR 14-AUG-2000; 2000US-0224518P.  
PR 14-AUG-2000; 2000US-0224519P.  
PR 14-AUG-2000; 2000US-0225213P.  
PR 14-AUG-2000; 2000US-0225214P.  
PR 14-AUG-2000; 2000US-0225266P.  
PR 14-AUG-2000; 2000US-0225267P.  
PR 14-AUG-2000; 2000US-0225268P.  
PR 14-AUG-2000; 2000US-0225270P.  
PR 14-AUG-2000; 2000US-0225477P.  
PR 14-AUG-2000; 2000US-0225757P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 14-AUG-2000; 2000US-0225759P.  
PR 22-AUG-2000; 2000US-0226681P.  
PR 22-AUG-2000; 2000US-0226682P.  
PR 22-AUG-2000; 2000US-0227182P.  
PR 23-AUG-2000; 2000US-0227009P.  
PR 30-AUG-2000; 2000US-0228924P.  
PR 01-SEP-2000; 2000US-0229287P.  
PR 01-SEP-2000; 2000US-0229343P.  
PR 01-SEP-2000; 2000US-0229344P.  
PR 01-SEP-2000; 2000US-0229345P.  
PR 05-SEP-2000; 2000US-0229509P.  
PR 05-SEP-2000; 2000US-0229513P.  
PR 06-SEP-2000; 2000US-0230337P.  
PR 06-SEP-2000; 2000US-0230338P.  
PR 08-SEP-2000; 2000US-0231242P.  
PR 08-SEP-2000; 2000US-0231243P.  
PR 08-SEP-2000; 2000US-0231244P.  
PR 08-SEP-2000; 2000US-0231413P.  
PR 08-SEP-2000; 2000US-0231414P.  
PR 08-SEP-2000; 2000US-0232080P.  
PR 08-SEP-2000; 2000US-0232081P.  
PR 12-SEP-2000; 2000US-0231968P.  
PR 14-SEP-2000; 2000US-0232397P.  
PR 14-SEP-2000; 2000US-0232398P.  
PR 14-SEP-2000; 2000US-0232399P.  
PR 14-SEP-2000; 2000US-0232400P.  
PR 14-SEP-2000; 2000US-0232401P.  
PR 14-SEP-2000; 2000US-0233063P.  
PR 14-SEP-2000; 2000US-0233064P.  
PR 14-SEP-2000; 2000US-0233065P.  
PR 21-SEP-2000; 2000US-0234223P.  
PR 21-SEP-2000; 2000US-0234274P.  
PR 25-SEP-2000; 2000US-0234997P.  
PR 25-SEP-2000; 2000US-0234998P.  
PR 26-SEP-2000; 2000US-0235884P.  
PR 27-SEP-2000; 2000US-0235884P.  
PR 27-SEP-2000; 2000US-0235834P.  
PR 29-SEP-2000; 2000US-0235836P.  
PR 29-SEP-2000; 2000US-0236327P.  
PR 29-SEP-2000; 2000US-0236367P.  
PR 29-SEP-2000; 2000US-0236368P.  
PR 29-SEP-2000; 2000US-0236369P.  
PR 29-SEP-2000; 2000US-0236370P.  
PR 02-OCT-2000; 2000US-0236802P.  
PR 02-OCT-2000; 2000US-0237037P.  
PR 02-OCT-2000; 2000US-0237038P.  
PR 02-OCT-2000; 2000US-0237039P.  
PR 02-OCT-2000; 2000US-0237040P.  
PR 13-OCT-2000; 2000US-0239335P.  
PR 13-OCT-2000; 2000US-0239337P.  
PR 20-OCT-2000; 2000US-0240960P.  
PR 20-OCT-2000; 2000US-0241221P.  
PR 20-OCT-2000; 2000US-0241785P.



QY 205 TVRQVSTLLXMCCHGNPAQYERNR 231  
 DB 1 TVRQVSTLLXMCCHGNPAQYERXR 27

## RESULT 12

ID ADC95956 standard; protein; 123 AA.

AC ADC95956;

DT 01-JAN-2004 (first entry)

DE E. faecium protein sequence SEQ ID 5583.

KW Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;  
 abdominal-pelvic infection.

OS Enterococcus faecium.

PN US6583275-B1.

PD 24-JUN-2003.

PF 30-JUN-1998; 98US-00107532.

PR 02-JUL-1997; 97US-0051571P.

PR 14-MAY-1998; 98US-0085598P.

PA (GENO-) GENOME THERAPEUTICS CORP.

PI Doucette-Stamm LA, Bush D;

DR WPI; 2003-799836/75.

DR N-PSDB; ADC92302.

PT New isolated nucleic acid derived from Enterococcus faecium encoding an  
 Enterococcus faecium polypeptide useful for detection, prevention and  
 treatment of a pathological condition resulting from a bacterial  
 infection.

PS Example 1; SEQ ID NO 5583; 243pp; English.

XX The invention relates to an isolated nucleic acid derived from  
 CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having  
 CC one of 10 fully defined sequences given in the (or comprising 40  
 CC sequential nucleotides chosen from any of the nucleic acids, its  
 CC complement or sequences hybridizing to it). Also included are a  
 CC recombinant vector comprising the nucleic acid operably linked to a  
 CC transcription regulatory element, a cell comprising the vector and a  
 CC single-stranded probe comprising the nucleic acid. The nucleic acids are  
 CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.  
 CC The nucleic acids is useful for diagnosing pathological conditions  
 CC resulting from E. faecium bacterial infection (e.g. urinary tract  
 CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic  
 CC infection) and for screening drugs such as agonists and antagonists. The  
 CC nucleic acid is useful for recombinant production of Candida albicans -  
 CC derived peptides or antisense polypeptides. Pharmaceutical compositions  
 CC and vaccines containing the nucleic acid are useful for preventing or  
 CC treating Enterococcus faecium infections. The present sequence represents  
 CC one if the disclosed E. faecium proteins.

CC Sequence 123 AA;

Query Match 3.3%; Score 122; DB 7; Length 123;  
 Best Local Similarity 35.2%; Pred. No. 6e-05;  
 Matches 37; Conservative 6; Mismatches 62; Indels 0; Gaps 0;

QY 105 GELREDNLPYKRAKARLLIFSTNTDXESGARSXPFGFAXXVKKYTTGTTGUMRP 164  
 DB 8 GELRWGNHSHSVKAEGLSLTARPTSRAGTKGLSDPVVPHGRALAQRIKATLIGTGLSP 67  
 QY 165 SXHSDVAFKSPDVGSYHKEAFTKRWIVHPLIGXSMVTVVRQ 209

DB 68 RVHIDSEVWHLIDVSSHPGAVGPKGMAYRPLRRHNSWQNVVRQ 112

## RESULT 13

ID ADC95957 standard; protein; 123 AA.

AC ADC95957;

DT 01-JAN-2004 (first entry)

DE E. faecium protein sequence SEQ ID 5584.

KW Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;  
 abdominal-pelvic infection.

OS Enterococcus faecium.

PN US6583275-B1.

PD 24-JUN-2003.

PF 30-JUN-1998; 98US-00107532.

PR 02-JUL-1997; 97US-0051571P.

PR 14-MAY-1998; 98US-0085598P.

PA (GENO-) GENOME THERAPEUTICS CORP.

PI Doucette-Stamm LA, Bush D;

DR WPI; 2003-799836/75.

DR N-PSDB; ADC92303.

PT New isolated nucleic acid derived from Enterococcus faecium encoding an  
 Enterococcus faecium polypeptide useful for detection, prevention and  
 treatment of a pathological condition resulting from a bacterial  
 infection.

PS Example 1; SEQ ID NO 5584; 243pp; English.

XX The invention relates to an isolated nucleic acid derived from  
 CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having  
 CC one of 10 fully defined sequences given in the (or comprising 40  
 CC sequential nucleotides chosen from any of the nucleic acids, its  
 CC complement or sequences hybridizing to it). Also included are a  
 CC recombinant vector comprising the nucleic acid operably linked to a  
 CC transcription regulatory element, a cell comprising the vector and a  
 CC single-stranded probe comprising the nucleic acid. The nucleic acids are  
 CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.  
 CC The nucleic acids is useful for diagnosing pathological conditions  
 CC resulting from E. faecium bacterial infection (e.g. urinary tract  
 CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic  
 CC infection) and for screening drugs such as agonists and antagonists. The  
 CC nucleic acid is useful for recombinant production of Candida albicans -  
 CC derived peptides or antisense polypeptides. Pharmaceutical compositions  
 CC and vaccines containing the nucleic acid are useful for preventing or  
 CC treating Enterococcus faecium infections. The present sequence represents  
 CC one if the disclosed E. faecium proteins.

CC Sequence 123 AA;

Query Match 3.3%; Score 122; DB 7; Length 123;  
 Best Local Similarity 35.2%; Pred. No. 6e-05;  
 Matches 37; Conservative 6; Mismatches 62; Indels 0; Gaps 0;

QY 105 GELREDNLPYKRAKARLLIFSTNTDXESGARSXPFGFAXXVKKYTTGTTGUMRP 164  
 DB 8 GELRWGNHSHSVKAEGLSLTARPTSRAGTKGLSDPVVPHGRALAQRIKATLIGTGLSP 67  
 QY 165 SXHSDVAFKSPDVGSYHKEAFTKRWIVHPLIGXSMVTVVRQ 209





PR	24-FEB-2000	2000US-01846664P
PR	02-MAR-2000	2000US-01983501P
PR	16-MAR-2000	2000US-0198374P
PR	17-MAR-2000	2000US-0198076P
PR	16-APR-2000	2000US-01981123P
PR	19-MAY-2000	2000US-0205515P
PR	07-JUN-2000	2000US-0205467P
PR	28-JUN-2000	2000US-0214886P
PR	30-JUN-2000	2000US-0215135P
PR	07-JUL-2000	2000US-0216647P
PR	07-JUL-2000	2000US-0216800P
PR	11-JUL-2000	2000US-0217487P
PR	11-JUL-2000	2000US-0217496P
PR	11-JUL-2000	2000US-0218290P
PR	26-JUL-2000	2000US-0218962P
PR	26-JUL-2000	2000US-0220964P
PR	14-AUG-2000	2000US-0224518P
PR	14-AUG-2000	2000US-0224519P
PR	14-AUG-2000	2000US-0225133P
PR	14-AUG-2000	2000US-0225143P
PR	14-AUG-2000	2000US-0225578P
PR	14-AUG-2000	2000US-0225579P
PR	18-AUG-2000	2000US-0226479P
PR	22-AUG-2000	2000US-0226681P
PR	22-AUG-2000	2000US-0226686P
PR	22-AUG-2000	2000US-0227182P
PR	30-AUG-2000	2000US-0227009P
PR	23-SEP-2000	2000US-0228924P
PR	01-SEP-2000	2000US-0229287P
PR	01-SEP-2000	2000US-0229343P
PR	01-SEP-2000	2000US-0229344P
PR	01-SEP-2000	2000US-0229345P
PR	01-SEP-2000	2000US-0229509P
PR	05-SEP-2000	2000US-0229517P
PR	06-SEP-2000	2000US-0230437P
PR	06-SEP-2000	2000US-0230438P
PR	08-SEP-2000	2000US-0231342P
PR	08-SEP-2000	2000US-0231343P
PR	08-SEP-2000	2000US-0231344P
PR	08-SEP-2000	2000US-0231345P
PR	08-SEP-2000	2000US-0231143P
PR	08-SEP-2000	2000US-0231144P
PR	08-SEP-2000	2000US-0231808P
PR	08-SEP-2000	2000US-0231809P
PR	12-SEP-2000	2000US-0231868P
PR	14-SEP-2000	2000US-0232397P
PR	14-SEP-2000	2000US-0232398P
PR	14-SEP-2000	2000US-0232399P
PR	14-SEP-2000	2000US-0232400P
PR	14-SEP-2000	2000US-0232401P
PR	14-SEP-2000	2000US-0233063P
PR	14-SEP-2000	2000US-0233064P
PR	14-SEP-2000	2000US-0233065P
PR	21-SEP-2000	2000US-0233423P
PR	21-SEP-2000	2000US-0234274P
PR	21-SEP-2000	2000US-0234997P
PR	25-SEP-2000	2000US-0234988P
PR	25-SEP-2000	2000US-0235484P
PR	26-SEP-2000	2000US-0235834P
PR	27-SEP-2000	2000US-0235834P
PR	27-SEP-2000	2000US-0235835P
PR	29-SEP-2000	2000US-0236337P
PR	29-SEP-2000	2000US-0236337P
PR	29-SEP-2000	2000US-0236338P
PR	29-SEP-2000	2000US-0236359P
PR	29-SEP-2000	2000US-0236370P
PR	02-OCT-2000	2000US-0237602P
PR	02-OCT-2000	2000US-0237603P
PR	02-OCT-2000	2000US-0237603P



Query Match 11.6%; Score 435; DB 4; Length 538;  
 Best Local Similarity 28.8%; Pred. No. 6.9e-40;  
 Matches 182; Conservative 50; Mismatches 216; Indels 184; Gaps 30;

123 LILISTNDXESGASRFPGFAXXVKTGTGTLGRPSXSDVAFXSFVGSYH 182  
 1 MILISTNDRSGAS-----DLTFVLSRRCQ-----KSYH 33  
 183 XEAFTKMTV-----HPLIGXSWXTVV-----RQVS----- 211  
 34 RD-----NMLVAARSKRRRFLILRCFLSLGSRHQLDCSPTRRERELGDRRETGL 88  
 212 FTLLMCCCHGNPQVYERNRKHLYVVLGGANGAKXXSVGLXNASKSRRPGTIRQ 271  
 89 FTLLMCCCHGNPQVYERNRKHLYVVLGGANGAKLPVGLLXNASKSRRPGTIRQ 148  
 272 RGASVGLGXPRPLSPAGRPPTSTXKXAGRVPRAPGGSXAECSSMETGRGKG 331  
 149 RGASVGLGWVPRPLSPAGAPPP-----RAP-----RAREG 181  
 332 GXPLARH-----APHVRAAEFYXSSSTINRHHSACIFMXXXILFLWVDIOXWDCX 382  
 182 ACPAARRDRGVRSLRPGKGAAGKAAPSVTHTFTVLVPSASAVAPTIDIOV-HVS 240  
 383 XTXMFYFWF-----IEKSYXXXXXLYKFTSLPSDPFKEXKRWVRNAPHXYP-PFTYL 434  
 241 LLYNEFFSGEISSGSIASPGSSISGLR--NLHTDFH-----RGCTNLHPYQVIFL 291  
 435 LQDFKGRYLOVSKXQIEYXNFCIRGT--DFXIOSCMNDKSRDLSXNWSQMKY 491  
 292 NMEWGEWML-ISTPLPLCYCRMTSRGTGYKHYTDKLNIEISALEQIIGFK----- 345  
 492 ISSSTXSGST-----ELAKSSLI--PTY-----XXXXKGFISNLIKG 528  
 346 VVTRTSLAGTCAGIGNLRNTFLVPRAYILLNMLCDHLYLLIKTNEKGLYQIYFYK 405  
 529 IKIKXXVKLFSLAFFONIKXXPSIGHLV-----CTRHCVCHXSKMFSXGSG 576  
 406 KKSNYRKSCLLHFNFKTKSYALLG-IYVDPDVFTFOKCS-HGVVNNNSV-GH----- 458  
 577 XFCRVXSLTVVLFKSKNLTKXXFNLRKXSNRTRTXIXITLQISPYHTAS--TCACKL 634  
 459 -----RNSLTVRLF--RNETSEPM-----SSTCAKLVTLGGLPEPSSHSRSLHTTLAH 506  
 635 IPGSCYF-PFYFLSLXLTTPSPHFFSGFLIV 665  
 507 VPVILFLAPLFLPPSSSHNPPFPPL-FLIFL 537

RESULT 2  
 US-09-643-476-2  
 Sequence 2, Application US/09643476  
 Patent No. 6599495

GENERAL INFORMATION:  
 APPLICANT: Evlin, Jr., Paul R.  
 TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCE  
 OF MAMMARY TISSUE AND METHODS OF USE  
 NUMBER OF SEQUENCES: 8  
 CORRESPONDENCE ADDRESS:  
 ADDRESS: Merchant & Gould P.C.  
 STREET: P.O. Box 2903  
 CITY: Minneapolis  
 STATE: MN  
 COUNTRY: USA

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette  
 COMPUTER: IBM Compatible  
 OPERATING SYSTEM: DOS  
 SOFTWARE: FASTSEQ for Windows Version 2.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/643,476  
 FILING DATE: 22-Aug-2000

## CLASSIFICATION: &lt;Unknown&gt;

## PRIOR APPLICATION DATA:

Application Number: 60/027,315  
 Filing Date: 03-OCT-1996  
 Application Number: PCT/US97/18026  
 Filing Date: 03-OCT-1997  
 Application Number: 08/943,828  
 Filing Date: 03-OCT-1997  
 Application Number: 08/994,076  
 Filing Date: 19-DEC-1997

## ATTORNEY/AGENT INFORMATION:

Name: Weaver, Katie G.  
 Registration Number: 43,245  
 Reference/Docket Number: 4273.1USC2  
 TELECOMMUNICATION INFORMATION:  
 Telephone: 612-332-5300  
 Telefax: 612-332-9081

## INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 538 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 FRAGMENT TYPE: internal  
 SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-643-476-2

Query Match 11.6%; Score 435; DB 4; Length 538;  
 Best Local Similarity 28.8%; Pred. No. 6.9e-40;  
 Matches 182; Conservative 50; Mismatches 216; Indels 184; Gaps 30;

123 LILISTNDXESGASRFPGFAXXVKTGTGTLGRPSXSDVAFXSFVGSYH 182  
 1 MILISTNDRSGAS-----DLTFVLSRRCQ-----KSYH 33  
 183 XEAFTKMTV-----HPLIGXSWXTVV-----RQVS----- 211  
 34 RD-----NMLVAARSKRRRFLILRCFLSLGSRHQLDCSPTRRERELGDRRETGL 88  
 212 FTLLMCCCHGNPQVYERNRKHLYVVLGGANGAKXXSVGLXNASKSRRPGTIRQ 271  
 89 FTLLMCCCHGNPQVYERNRKHLYVVLGGANGAKLPVGLLXNASKSRRPGTIRQ 148  
 272 RGASVGLGXPRPLSPAGRPPTSTXKXAGRVPRAPGGSXAECSSMETGRGKG 331  
 149 RGASVGLGWVPRPLSPAGAPPP-----RAP-----RAREG 181  
 332 GXPLARH-----APHVRAAEFYXSSSTINRHHSACIFMXXXILFLWVDIOXWDCX 382  
 182 ACPAARRDRGVRSLRPGKGAAGKAAPSVTHTFTVLVPSASAVAPTIDIOV-HVS 240  
 383 XTXMFYFWF-----IEKSYXXXXXLYKFTSLPSDPFKEXKRWVRNAPHXYP-PFTYL 434  
 241 LLYNEFFSGEISSGSIASPGSSISGLR--NLHTDFH-----RGCTNLHPYQVIFL 291  
 435 LQDFKGRYLOVSKXQIEYXNFCIRGT--DFXIOSCMNDKSRDLSXNWSQMKY 491  
 292 NMEWGEWML-ISTPLPLCYCRMTSRGTGYKHYTDKLNIEISALEQIIGFK----- 345  
 492 ISSSTXSGST-----ELAKSSLI--PTY-----XXXXKGFISNLIKG 528  
 346 VVTRTSLAGTCAGIGNLRNTFLVPRAYILLNMLCDHLYLLIKTNEKGLYQIYFYK 405  
 529 IKIKXXVKLFSLAFFONIKXXPSIGHLV-----CTRHCVCHXSKMFSXGSG 576  
 406 KKSNYRKSCLLHFNFKTKSYALLG-IYVDPDVFTFOKCS-HGVVNNNSV-GH----- 458  
 577 XFCRVXSLTVVLFKSKNLTKXXFNLRKXSNRTRTXIXITLQISPYHTAS--TCACKL 634  
 459 -----RNSLTVRLF--RNETSEPM-----SSTCAKLVTLGGLPEPSSHSRSLHTTLAH 506

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2004, 10:51:05 / Search time 19.8859 Seconds  
(without alignments)  
2035.354 Million cell updates/sec

Title: US-10-028-952A-10

Perfect score: 3752

Sequence: 1 IRHKGHETXEVXNKKKAPG.....DXALALQPKXRETTLSQKK 784

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Issued Patents AA:  
1: /cgn2.6/prodata/2/1aa/5A\_COMB.pep.\*  
2: /cgn2.6/prodata/2/1aa/5B\_COMB.pep.\*  
3: /cgn2.6/prodata/2/1aa/5A\_COMB.pep.\*  
4: /cgn2.6/prodata/2/1aa/5B\_COMB.pep.\*  
5: /cgn2.6/prodata/2/1aa/5A\_COMB.pep.\*  
6: /cgn2.6/prodata/2/1aa/5B\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	435	11.6	538	US-08-994-076-2	Sequence 2, Appli
2	435	11.6	538	US-09-643-476-2	Sequence 2, Appli
3	133.5	3.6	497	US-09-252-991A-22620	Sequence 23620, A
4	122.5	3.3	67	US-09-621-976-394	Sequence 394, A
5	122	3.3	123	US-09-107-532A-5583	Sequence 5583, A
6	122	3.3	123	US-09-107-532A-5584	Sequence 5584, A
7	122	3.3	123	US-09-134-000C-6555	Sequence 6555, A
8	120.5	3.2	555	US-09-252-991A-28848	Sequence 28848, A
9	119	3.2	97	US-09-134-001C-3072	Sequence 3072, A
10	119	3.2	97	US-09-134-001C-3726	Sequence 3726, A
11	119	3.2	97	US-09-134-001C-3875	Sequence 3875, A
12	119	3.2	97	US-09-134-001C-4192	Sequence 4192, A
13	119	3.2	97	US-09-134-001C-4844	Sequence 4844, A
14	117	3.1	106	US-09-134-001C-3312	Sequence 3312, A
15	116.5	3.1	456	US-09-252-991A-17335	Sequence 17335, A
16	116	3.1	82	US-09-134-000C-4749	Sequence 4749, A
17	115.5	3.1	653	US-09-252-991A-21591	Sequence 21591, A
18	115	3.1	405	US-09-252-991A-24483	Sequence 24483, A
19	115	3.1	425	US-09-252-991A-20467	Sequence 20467, A
20	114.5	3.1	728	US-09-252-991A-22613	Sequence 22613, A
21	113.5	3.0	719	US-09-252-991A-23660	Sequence 23660, A
22	113	3.0	568	US-09-252-991A-23264	Sequence 23264, A
23	112.5	3.0	390	US-09-252-991A-33044	Sequence 33044, A
24	111	3.0	397	US-09-252-991A-24025	Sequence 24025, A
25	110.5	2.9	1020	US-09-252-991A-28870	Sequence 28870, A
26	110	2.9	348	US-09-252-991A-25605	Sequence 25605, A
27	109.5	2.9	1113	US-09-252-991A-29215	Sequence 29215, A

28	108.5	2.9	312	US-09-252-991A-20216	Sequence 20216, A
29	108	2.9	274	US-09-252-991A-28347	Sequence 28347, A
30	108	2.9	323	US-09-252-991A-23885	Sequence 23885, A
31	108	2.9	668	US-09-252-991A-22794	Sequence 22794, A
32	107	2.9	496	US-09-252-991A-25366	Sequence 25366, A
33	106.5	2.8	309	US-09-252-991A-25386	Sequence 25386, A
34	106.5	2.8	472	US-09-252-991A-31978	Sequence 31978, A
35	106	2.8	70	US-09-621-976-4551	Sequence 4551, A
36	106	2.8	530	US-09-252-991A-20846	Sequence 20846, A
37	105.5	2.8	485	US-09-252-991A-22450	Sequence 22450, A
38	105	2.8	260	US-09-252-991A-28669	Sequence 28669, A
39	105	2.8	628	US-09-252-991A-32727	Sequence 32727, A
40	104.5	2.8	478	US-09-252-991A-24169	Sequence 24169, A
41	104	2.8	693	US-09-252-991A-26071	Sequence 26071, A
42	103	2.7	603	US-09-252-991A-22674	Sequence 22674, A
43	102.5	2.7	354	US-09-252-991A-26135	Sequence 26135, A
44	102.5	2.7	310	US-09-252-991A-19986	Sequence 19986, A
45	102.5	2.7	606	US-09-252-991A-18233	Sequence 18233, A

## ALIGNMENTS

## RESULT 1

US-08-994-076-2  
Sequence 2, Application US/08994076

Patent No. 650937

GENERAL INFORMATION:

APPLICANT: Evin, Jr., Paul R.

TITLE OF INVENTION: NUCLEOTIDE AND PROTEIN SEQUENCE

TITLE OF INVENTION: OF MAMMASTATIN AND METHODS OF USE

NUMBER OF SEQUENCES: 8

CORRESPONDENCE ADDRESS:

ADDRESSER: Merchant, Gould, Smith, Edell, Welter & Schmidt

STREET: 3100 No. 6500937west Center, 90 South 7th Street

CITY: Minneapolis

STATE: MN

COUNTRY: USA

ZIP: 55402

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM Compatible

OPERATING SYSTEM: DOS

SOFTWARE: FastSeq for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/994,076

FILING DATE: 19-DEC-1997

CLASSIFICATION:

Prior Application Data:

APPLICATION NUMBER: 60/027,315

FILING DATE: 03-OCT-1996

APPLICATION NUMBER: PCT/US97/18026

FILING DATE: 03-OCT-1997

APPLICATION NUMBER: 08/943,828

ATTORNEY/AGENT INFORMATION:  
NAME: Kettelberger, Denise X  
REGISTRATION NUMBER: 33,924  
REFERENCE/DOCKET NUMBER: 4273,1US11  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 612-332-5300  
TELEFAX: 612-332-9081  
TELEX:  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 538 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
FRAGMENT TYPE: internal  
US-08-994-076-2

QY 635 IPGSCVF-PFYFLSLXTTTPEFHPFFSFLIV 665  
 DB 507 VPVILFLAPXLFPPFSHPNPFPPFL-FLFLP 537

# RESULT 3

US-09-252-991A-23620  
 ; Sequence 23620, Application US/09252991A  
 ; Patent No. 6551795

## GENERAL INFORMATION:

APPLICANT: Mate J. Rubenfield et al.  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
 FILE REFERENCE: 107196.136  
 CURRENT APPLICATION NUMBER: US/09/252,991A  
 CURRENT FILING DATE: 1999-02-18  
 PRIOR APPLICATION NUMBER: US 60/074,768  
 PRIOR FILING DATE: 1998-02-18  
 PRIOR APPLICATION NUMBER: US 60/094,190  
 PRIOR FILING DATE: 1998-07-27  
 NUMBER OF SEQ ID NOS: 33142  
 SEQ ID NO: 23620  
 LENGTH: 497  
 TYPE: PRT  
 ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-23620

Query Match 3.6%; Score 133.5; DB 4; Length 497;  
 Best Local Similarity 22.7%; Pred. No. 3.3e-06;  
 Matches 97; Conservative 30; Mismatches 135; Indels 165; Gaps 20;

QY 18 AGGAPPVSPRGA-----RGXRRPCPPVKNHSDXXTPVRRGGER-- 61  
 DB 112 AAPAAPARRRALLAQFRQVRPATAREGGRSPDPER-----RLRRGPLRRGQHPFR 165  
 QY 62 ----GALAXGAKRPA--RRPGATRSQGSARMGVXLGRYTCQTXOV-----SXGE 106  
 DB 166 RAIPGAAAPGRPRPALPRRYRPPRRRAGAPR--RMGQRLRRGALLAAGAAGATGA 222  
 QY 107 IREDN-----LPMXR-----PAKARLLITSTVDXESG--- 136  
 DB 223 VRTTRPVGVRPRLRRTARPGOPPGPLPARQGRPGKAGSDLRAPAPAGRAGDRO 282  
 QY 137 ---ASRSXPXPF-----GFXAXXVXKVTGIGT-----LWPSXHSVAFXSFDVGSSY 181  
 DB 283 GGRRLPAAGPGLPQTPGILHARRORRVAADGDSVRAPAWGGHADL----- 331  
 QY 182 HXAEFTKRWIVHPLIGXXSWYTVRQVSTLLXMCCHGNDPAQYENRREXHLVYLG 241  
 DB 332 --PGQPQARQLAEPGTG-----PAPARRQSRLRHLHLRLH 364  
 QY 242 XANANAKXXSVGLXNASKESRPXTTROR-----GAS-----VGLGKXPRLSPPA 290  
 DB 365 RPAEGGQHPRG---AGRAPMDAGHLPRRRRRADAEAGELRRVG---VLLAA 415  
 QY 291 GRPPSTRXKRAAGVPRAPPGGSAECPSMETGRGKGGPLARAPVPRAPAEFXX 350  
 DB 416 GHRLP-----AGRRPRAPRSGA-----PGGTGAIVGDHALRA--AA 453  
 QY 351 SSTINR 357  
 DB 454 AATVHR 460

# RESULT 4

US-09-621-976-3994  
 ; Sequence 3994, Application US/09621976  
 ; Patent No. 6639083

## GENERAL INFORMATION:

APPLICANT: Dumas Milne Edwards, J.B.  
 APPLICANT: Jober, S.  
 APPLICANT: Giordano, J.Y.  
 TITLE OF INVENTION: ESTs and Encoded Human Proteins.

FILE REFERENCE: GENSET.054PR2  
 CURRENT APPLICATION NUMBER: US/09/621,976  
 CURRENT FILING DATE: 2000-07-21  
 NUMBER OF SEQ ID NOS: 19335  
 SOFTWARE: Patent.pm  
 SEQ ID NO 3994  
 LENGTH: 67  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: SIGNAL  
 LOCATION: 26..1  
 US-09-621-976-3994

Query Match 3.3%; Score 122.5; DB 4; Length 67;  
 Best Local Similarity 52.5%; Pred. No. 3.8e-06;  
 Matches 31; Conservative 5; Mismatches 22; Indels 1; Gaps 1;

QY 726 KTKIXPVLAHTCSISYRGXGXIDXAQEVAAVRYDXATALPGXERETLSQKK 784  
 DB 8 ENTKMLGVVCAQSPSGRGRRJAWTEVAVSR-DQATALPGQORETLSQKK 65

# RESULT 5

US-09-107-532A-5583  
 ; Sequence 5583, Application US/09107532A  
 ; Patent No. 6583275

## GENERAL INFORMATION:

APPLICANT: Lynn A Doucelte-Stamm and David Bush  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS  
 NUMBER OF SEQUENCES: 7310  
 CORRESPONDENCE ADDRESS:  
 ADDRESSER: GENOME THERAPEUTICS CORPORATION  
 STREET: 100 Beaver Street  
 CITY: Waltham  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02354

## COMPUTER READABLE FORM:

MEDIUM TYPE: CD-ROM ISO9660  
 COMPUTER: PC  
 OPERATING SYSTEM: <Unknown>  
 SOFTWARE: ASCII  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/107,532A  
 FILING DATE: 30-Jun-1998

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/085,598  
 FILING DATE: 14 May 1998  
 APPLICATION NUMBER: 60/051571  
 FILING DATE: July 2, 1997  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Ariniello, Pamela Denike  
 REGISTRATION NUMBER: 40,489  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (781)893-5007  
 TELEFAX: (781)893-8277  
 INFORMATION FOR SEQ ID NO: 5583:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 123 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHEICAL: YES  
 ORIGINAL SOURCE:  
 ORGANISM: Enterococcus faecium

## FEATURE:

NAME/KEY: misc\_feature  
 LOCATION: (B) LOCATION 1...123  
 SEQUENCE DESCRIPTION: SEQ ID NO: 5583:  
 US-09-107-532A-5583

Query Match 3.3%; Score 122; DB 4; Length 123;  
 Best Local Similarity 35.2%; Pred. No. 9.8e-06;  
 Matches 37; Conservative 6; Mismatches 62; Indels 0; Gaps 0;

QY 105 GELREDNLPWXRARAKRLLIFSTNTDXESGASRSPFGFAXXVXKVTGIGLMP 164  
 DB 8 GELRWGNHKSXVKAEGSLTARPTSRAGTKVGLSDPVVPHGRALAKRIKATIGTGLSP 67

QY 165 SXHSDVAFKSPDVGSSTHKEAFTKRWIVHPLIGXSWXTYVRQ 209  
 DB 68 RVHIDGEVWHLDVGSSHPGAVVGPKGMAVRFLKRHSAWQNVVRQ 112

RESULT 6  
 US-09-107-532A-5584  
 ; Sequence 5584, Application US/09107532A  
 ; Patent No. 6583275

GENERAL INFORMATION:  
 APPLICANT: Lynn A Doucette-Stamm and David Bush  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

NUMBER OF SEQUENCES: 7310  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: GENOME THERAPEUTICS CORPORATION

STREET: 100 Beaver Street  
 CITY: Waltham  
 STATE: Massachusetts  
 COUNTRY: USA  
 ZIP: 02154

COMPUTER READABLE FORM:  
 MEDIUM TYPE: CD-ROM ISO9660  
 COMPUTER: PC  
 OPERATING SYSTEM: <Unknown>  
 SOFTWARE: ASCII

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/107,532A  
 FILING DATE: 30-Jun-1998

PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: 60/085,598  
 FILING DATE: 14 May 1998  
 APPLICATION NUMBER: 60/051,571  
 FILING DATE: July 2, 1997

ATTORNEY/AGENT INFORMATION:  
 NAME: Ariniello, Pamela Deneka  
 REGISTRATION NUMBER: 40,489  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (781)893-5507  
 TELEFAX: (781)893-8277

INFORMATION FOR SEQ ID NO: 5584:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 123 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear

MOLECULE TYPE: protein  
 HYPOTHETICAL: YES  
 ORIGINAL SOURCE:  
 ORGANISM: Enterococcus faecium

FEATURES:  
 NAME/KEY: misc\_feature  
 LOCATION: (B) LOCATION 1...123  
 SEQUENCE DESCRIPTION: SEQ ID NO: 5584:  
 US-09-107-532A-5584

Query Match 3.3%; Score 122; DB 4; Length 123;  
 Best Local Similarity 35.2%; Pred. No. 9.8e-06;  
 Matches 37; Conservative 6; Mismatches 62; Indels 0; Gaps 0;

QY 105 GELREDNLPWXRARAKRLLIFSTNTDXESGASRSPFGFAXXVXKVTGIGLMP 164  
 DB 8 GELRWGNHKSXVKAEGSLTARPTSRAGTKVGLSDPVVPHGRALAKRIKATIGTGLSP 67

QY 165 SXHSDVAFKSPDVGSSTHKEAFTKRWIVHPLIGXSWXTYVRQ 209  
 DB 68 RVHIDGEVWHLDVGSSHPGAVVGPKGMAVRFLKRHSAWQNVVRQ 112

RESULT 7  
 US-09-134-000C-6555  
 ; Sequence 6555, Application US/09134000C  
 ; Patent No. 6617156

GENERAL INFORMATION:  
 APPLICANT: Lynn Doucette-Stamm et al  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
 ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 032796-032  
 CURRENT APPLICATION NUMBER: US/09/134,000C

CURRENT FILING DATE: 1998-08-13  
 PRIOR APPLICATION NUMBER: US 60/055,778

PRIOR FILING DATE: 1997-08-15  
 NUMBER OF SEQ ID NOS: 6812

SOFTWARE: PatentIn version 3.1  
 SEQ ID NO 6555

LENGTH: 123  
 TYPE: PRT  
 ORGANISM: Enterococcus faecalis  
 US-09-134-000C-6555

Query Match 3.3%; Score 122; DB 4; Length 123;  
 Best Local Similarity 35.2%; Pred. No. 9.8e-06;  
 Matches 37; Conservative 6; Mismatches 62; Indels 0; Gaps 0;

QY 105 GELREDNLPWXRARAKRLLIFSTNTDXESGASRSPFGFAXXVXKVTGIGLMP 164  
 DB 8 GELRWGNHKSXVKAEGSLTARPTSRAGTKVGLSDPVVPHGRALAKRIKATIGTGLSP 67

QY 165 SXHSDVAFKSPDVGSSTHKEAFTKRWIVHPLIGXSWXTYVRQ 209  
 DB 68 RVHIDGEVWHLDVGSSHPGAVVGPKGMAVRFLKRHSAWQNVVRQ 112

RESULT 8  
 US-09-252-991A-28848  
 ; Sequence 28848, Application US/09252991A  
 ; Patent No. 6551795

GENERAL INFORMATION:  
 APPLICANT: Marc J. Rubenfield et al.  
 TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
 TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196-136  
 CURRENT APPLICATION NUMBER: US/09/252,991A

CURRENT FILING DATE: 1999-02-18  
 PRIOR APPLICATION NUMBER: US 60/074,788

PRIOR FILING DATE: 1998-02-18  
 PRIOR APPLICATION NUMBER: US 60/094,190

PRIOR FILING DATE: 1998-07-27  
 NUMBER OF SEQ ID NOS: 33142

SEQ ID NO 28848  
 LENGTH: 555

TYPE: PRT  
 ORGANISM: Pseudomonas aeruginosa  
 US-09-252-991A-28848

Query Match 3.2%; Score 120.5; DB 4; Length 555;  
 Best Local Similarity 22.5%; Pred. No. 0.00011;  
 Matches 93; Conservative 21; Mismatches 144; Indels 155; Gaps 19;

QY 15 KKEAPGAPVSP-----RGARGGRRPCGPPVTKYHSDRXKXTDVR-RGGEPRGALA 65  
 DB 83 REHCPHRAKPDSDHPRRWGRLGCGRRP--PTDRLEAAGBEERHPRRLRSAPAGARC 140

QY 66 XG-----AKRPARRP--GATRSXSARNGVYLGRTCTQTXQVSGX-----ELR 108  
 DB 141 FGPRORPGKPDADADAPRRDVRGVRAGQ---GRANHGCHRLRLRRVAGGAPLEAR 196



```
QY 109 EDENLFWKRRK---ARLLIFSTNTDSESGASRSPFGFXAXY-----RKYT 155
; TYPE: PRY
; ORGANISM: Staphylococcus epidermidis
Db 197 PHRAAPGRLPRNRHAGTVGDLHDHRRPRAGSRVLMCGMAAPVOAARPDRTNR 256
QY 156 TGTGLMRPSXHSVAKSPVGVSSYHXEAFTRMIVHPLIGXSWXTVVRQVSTLL 215
; TYPE: PRY
; ORGANISM: Staphylococcus epidermidis
Db 257 LARAGALRRAGHLDPAG-----RRR----- 275
QY 216 XMCCCHGNPAQYEENRERXRLVYLGXGANGAKXXSVGLXLMXSKSESPXGTIRORGA 275
; TYPE: PRY
; ORGANISM: Staphylococcus epidermidis
Db 276 -----RPRPRVRRRCRNLPH-----RYRSAG--NAGARQAGNAGRGARAGV 316
QY 276 SVGLG-----XKXRL---SPP---AGRPPTRXKXRGAVPRRAG 312
; TYPE: PRY
; ORGANISM: Staphylococcus epidermidis
Db 317 GMSGGRASAGLRGTAGVPERDAPSPRADHPRSPHRLAGNPPRRDR--RTGRSLRRGQG 375
QY 313 PGSAE-----CPSSWETGRGRGXPLAHPHVRAR 345
; TYPE: PRY
; ORGANISM: Staphylococcus epidermidis
Db 376 QGRAGTAARSGRGRTARGGCLAGLCP-----RGARGRCHRRHRPQRRGR 422

RESULT 9
US-09-134-001C-3072
; Sequence 3072, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3072
; LENGTH: 97
; TYPE: PRY
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3072

Query Match
Best Local Similarity 3.2%; Score 119; DB 4; Length 97;
Matches 31; Conservative 4; Mismatches 47; Indels 0; Gaps 0;

QY 128 STNTDSESGASRSPFGFXAXYRKVTGTTGLMRPSXHSVDVAFSPDVGSYHXEA 187
; TYPE: PRY
; ORGANISM: Staphylococcus epidermidis
Db 8 TSRAKSGDGLSDPVVPHGRALAQRIKATPGITGLSPRVHIDGEVWHLDVSSHPGAVG 67
QY 188 TKRWIVHPLIGXSWXTVVRQ 209
; TYPE: PRY
; ORGANISM: Staphylococcus epidermidis
Db 68 PKGMAVRPLKRYASVQNVVRQ 89

RESULT 10
US-09-134-001C-3726
; Sequence 3726, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3726
; LENGTH: 97
; TYPE: PRY
; ORGANISM: Staphylococcus epidermidis
```

```
; LENGTH: 97
; TYPE: PRY
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3726

Query Match
Best Local Similarity 3.2%; Score 119; DB 4; Length 97;
Matches 31; Conservative 4; Mismatches 47; Indels 0; Gaps 0;

QY 128 STNTDSESGASRSPFGFXAXYRKVTGTTGLMRPSXHSVDVAFSPDVGSYHXEA 187
; TYPE: PRY
; ORGANISM: Staphylococcus epidermidis
Db 8 TSRAKSGDGLSDPVVPHGRALAQRIKATPGITGLSPRVHIDGEVWHLDVSSHPGAVG 67
QY 188 TKRWIVHPLIGXSWXTVVRQ 209
; TYPE: PRY
; ORGANISM: Staphylococcus epidermidis
Db 68 PKGMAVRPLKRYASVQNVVRQ 89

RESULT 11
US-09-134-001C-3875
; Sequence 3875, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 3875
; LENGTH: 97
; TYPE: PRY
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-3875

Query Match
Best Local Similarity 3.2%; Score 119; DB 4; Length 97;
Matches 31; Conservative 4; Mismatches 47; Indels 0; Gaps 0;

QY 128 STNTDSESGASRSPFGFXAXYRKVTGTTGLMRPSXHSVDVAFSPDVGSYHXEA 187
; TYPE: PRY
; ORGANISM: Staphylococcus epidermidis
Db 8 TSRAKSGDGLSDPVVPHGRALAQRIKATPGITGLSPRVHIDGEVWHLDVSSHPGAVG 67
QY 188 TKRWIVHPLIGXSWXTVVRQ 209
; TYPE: PRY
; ORGANISM: Staphylococcus epidermidis
Db 68 PKGMAVRPLKRYASVQNVVRQ 89

RESULT 12
US-09-134-001C-4192
; Sequence 4192, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
; FILE REFERENCE: GTC-007
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 4192
; LENGTH: 97
; TYPE: PRY
; ORGANISM: Staphylococcus epidermidis
```

US-09-134-001C-4192

Query Match 3.2%; Score 119; DB 4; Length 97;  
Best Local Similarity 37.8%; Pred. No. 1.6e-05;  
Matches 31; Conservative 4; Mismatches 47; Indels 0; Gaps 0;

QY 128 STNTDSESGASRSFPGFAXXXVRRKTTGTLMPKPSXHSVDVAFSPDVSSYHXKAEF 187  
DB 8 TSBAGSKDGLSDPVVPHGRALIAQRKATPGITGLSPRVHIDGVWHLVDVSSHPGAVVG 67

QY 188 TKRMTVHPLIGXXSMVXTTVRRQ 209  
DB 68 PKGMVAPLKRKYASWQNVVRRQ 89

RESULT 13

US-09-134-001C-4844  
Sequence 4844, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 4844  
LENGTH: 97  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-4844

Query Match 3.2%; Score 119; DB 4; Length 97;  
Best Local Similarity 37.8%; Pred. No. 1.6e-05;  
Matches 31; Conservative 4; Mismatches 47; Indels 0; Gaps 0;

QY 128 STNTDSESGASRSFPGFAXXXVRRKTTGTLMPKPSXHSVDVAFSPDVSSYHXKAEF 187  
DB 8 TSBAGSKDGLSDPVVPHGRALIAQRKATPGITGLSPRVHIDGVWHLVDVSSHPGAVVG 67  
QY 188 TKRMTVHPLIGXXSMVXTTVRRQ 209  
DB 68 PKGMVAPLKRKYASWQNVVRRQ 89

RESULT 14

US-09-134-001C-3312  
Sequence 3312, Application US/09134001C  
Patent No. 6380370  
GENERAL INFORMATION:  
APPLICANT: Lynn Doucette-Stamm et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: GTC-007  
CURRENT APPLICATION NUMBER: US/09/134,001C  
CURRENT FILING DATE: 1998-08-13  
PRIOR APPLICATION NUMBER: US 60/064,964  
PRIOR FILING DATE: 1997-11-08  
PRIOR APPLICATION NUMBER: US 60/055,779  
PRIOR FILING DATE: 1997-08-14  
NUMBER OF SEQ ID NOS: 5674  
SEQ ID NO 3312  
LENGTH: 106  
TYPE: PRT  
ORGANISM: Staphylococcus epidermidis  
US-09-134-001C-3312

Query Match 3.1%; Score 117; DB 4; Length 106;

Best Local Similarity 37.3%; Pred. No. 2.9e-05;  
Matches 31; Conservative 4; Mismatches 48; Indels 0; Gaps 0;

QY 128 STNTDSESGASRSFPGFAXXXVRRKTTGTLMPKPSXHSVDVAFSPDVSSYHXKAEF 187  
DB 8 TSBAGSKDGLSDPVVPHGRALIAQRKATPGITGLSPRVHIDGVWHLVDVSSHPGAVVG 67

QY 188 TKRMTVHPLIGXXSMVXTTVRRQ 210  
DB 68 PKGMVAPLKRKYASWQNVVRRQ 90

RESULT 15

US-09-252-991A-17335  
Sequence 17335, Application US/09252991A  
Patent No. 6551795  
GENERAL INFORMATION:  
APPLICANT: Marc J. Rubenfield et al  
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
FILE REFERENCE: 107196.136  
CURRENT APPLICATION NUMBER: US/09/252,991A  
CURRENT FILING DATE: 1999-02-18  
PRIOR APPLICATION NUMBER: US 60/074,788  
PRIOR FILING DATE: 1998-02-18  
PRIOR APPLICATION NUMBER: US 60/094,190  
PRIOR FILING DATE: 1998-07-27  
NUMBER OF SEQ ID NOS: 33142  
SEQ ID NO 17335  
LENGTH: 456  
TYPE: PRT  
ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-17335

Query Match 3.1%; Score 116.5; DB 4; Length 456;  
Best Local Similarity 23.0%; Pred. No. 0.00023;  
Matches 84; Conservative 13; Mismatches 126; Indels 141; Gaps 15;

QY 19 PGAPVSPRGARGGX-RRPCGPVYKH-YSDRXXTDPVRR-----GGEPRGALAXGAKRP 71  
DB 71 FGHVAAQPDGIRGCGVPRRPAAVHGRHRLRDAGCGFANHRRWFGRLPGRLLAATGQRP 130  
QY 72 AARRPGATRGXGASRMGVXIGRYTCQTXOVSXGELRED-----NLPKXRAKARLLILI 126  
DB 131 GGRPRGA-----GURTGAY-----LRGHAADPAARRRHHARRAARPAALV 174  
QY 127 -----FSTNTDSESGAS-RSEFXPGFAXXXVRRKTTGTLMPR 164  
DB 175 GHLAGGIRDARRHRHPRFALRDARPGQGVRRPQAGIPPGVRR----- 223  
QY 165 SXHSDVAFKSPDVSSYHXKAEFTKRWTVHPLIGXXSMVXTTVRRQVGSFTLLXMCCHGNP 224  
DB 224 -----ARFAG-----HG-- 230  
QY 225 AQYERNRXRLVTVLXGANGAKXXSVGLXLNASKESREXGTIRORRGASVGLKXPX 284  
DB 231 ---GPRR-----GAARAGHRRGVGLGQAASATPALPT-----RRAARFPGDPGQ 274  
QY 285 RLSPRAGPPPTSTXXKRAAGV--PRAPRGSALECESSWETGRK-----GX 333  
DB 275 PLAGAGRAAGPGTQPRGHRLATGPRAGGRSLPAAAGGGGGLHRCDDPALGGGGG 334  
QY 334 PLARHA 339  
DB 335 PARRPA 340

Search completed: April 27, 2004, 10:56:54  
Job time : 20.8859 secs



```

OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (249)..(250)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (255)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (266)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (281)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (283)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (299)..(300)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (316)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (333)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (349)..(350)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (366)..(368)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (378)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (382)..(383)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (386)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (398)..(401)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (416)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (427)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (433)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (449)..(450)
OTHER INFORMATION: Xaa is any amino acid or may be absent

```

```

FEATURE:
NAME/KEY: UNSURE
LOCATION: (456)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (466)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (483)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (499)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE

```

Query Match 94.5%; Score 3546; DB 14; Length 784;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 784; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 IRHEGEETXEVXNCKEAPGAPVSPRGARGGXRRPGCPVYKTHYSDRXXTDPVRRGGE 60
DB 1 IRHEGEETXEVXNCKEAPGAPVSPRGARGGXRRPGCPVYKTHYSDRXXTDPVRRGGE 60
QY 61 RGALXGAKRPAARRPGATRSXGSAWGVXIGRYTCOTXOVXGELREDENLPMXBRAX 120
DB 61 RGALXGAKRPAARRPGATRSXGSAWGVXIGRYTCOTXOVXGELREDENLPMXBRAX 120
QY 61 RGALXGAKRPAARRPGATRSXGSAWGVXIGRYTCOTXOVXGELREDENLPMXBRAX 120
DB 61 RGALXGAKRPAARRPGATRSXGSAWGVXIGRYTCOTXOVXGELREDENLPMXBRAX 120
QY 121 ARLLIFSTNTDXESGASRSFPFGFXAXVRYKTTGITGLMRPSXHSDDVAFXSFDVGS 180
DB 121 ARLLIFSTNTDXESGASRSFPFGFXAXVRYKTTGITGLMRPSXHSDDVAFXSFDVGS 180
QY 121 ARLLIFSTNTDXESGASRSFPFGFXAXVRYKTTGITGLMRPSXHSDDVAFXSFDVGS 180
DB 121 ARLLIFSTNTDXESGASRSFPFGFXAXVRYKTTGITGLMRPSXHSDDVAFXSFDVGS 180
QY 181 YHKEEFTKRWIVHPLIGXSMVTVVROVSFTLLMCCGNGNAQVRRRNRXHLVYL 240
DB 181 YHKEEFTKRWIVHPLIGXSMVTVVROVSFTLLMCCGNGNAQVRRRNRXHLVYL 240
QY 181 YHKEEFTKRWIVHPLIGXSMVTVVROVSFTLLMCCGNGNAQVRRRNRXHLVYL 240
DB 181 YHKEEFTKRWIVHPLIGXSMVTVVROVSFTLLMCCGNGNAQVRRRNRXHLVYL 240
QY 241 GXGANGAKXXSVGLXNLNASKESREPXGIRORRASVGLGXPRLSPPAGRPSTRX 300
DB 241 GXGANGAKXXSVGLXNLNASKESREPXGIRORRASVGLGXPRLSPPAGRPSTRX 300
QY 241 GXGANGAKXXSVGLXNLNASKESREPXGIRORRASVGLGXPRLSPPAGRPSTRX 300
DB 241 GXGANGAKXXSVGLXNLNASKESREPXGIRORRASVGLGXPRLSPPAGRPSTRX 300
QY 301 RAGGVVPRARPGSGXACBSSWETGRGRKCGXPLARAPVRRARAEYXSSSTIHNRHTS 360
DB 301 RAGGVVPRARPGSGXACBSSWETGRGRKCGXPLARAPVRRARAEYXSSSTIHNRHTS 360
QY 301 RAGGVVPRARPGSGXACBSSWETGRGRKCGXPLARAPVRRARAEYXSSSTIHNRHTS 360
DB 301 RAGGVVPRARPGSGXACBSSWETGRGRKCGXPLARAPVRRARAEYXSSSTIHNRHTS 360
QY 361 ACIFPXXXXLFLMWDIQMDCKXTWYFFWFIEKSSYXXXXRLTKFTSLPDPFKXBRMV 420
DB 361 ACIFPXXXXLFLMWDIQMDCKXTWYFFWFIEKSSYXXXXRLTKFTSLPDPFKXBRMV 420
QY 361 ACIFPXXXXLFLMWDIQMDCKXTWYFFWFIEKSSYXXXXRLTKFTSLPDPFKXBRMV 420
DB 361 ACIFPXXXXLFLMWDIQMDCKXTWYFFWFIEKSSYXXXXRLTKFTSLPDPFKXBRMV 420
QY 421 WRNAPHXYPPTXLIQNDFKGYRILQVSSXROIEYXNFCIRGTDPTXIQSCNNKDKCGRDL 480
DB 421 WRNAPHXYPPTXLIQNDFKGYRILQVSSXROIEYXNFCIRGTDPTXIQSCNNKDKCGRDL 480
QY 421 WRNAPHXYPPTXLIQNDFKGYRILQVSSXROIEYXNFCIRGTDPTXIQSCNNKDKCGRDL 480
DB 421 WRNAPHXYPPTXLIQNDFKGYRILQVSSXROIEYXNFCIRGTDPTXIQSCNNKDKCGRDL 480
QY 481 OSXWKSOMKXISSSTTSXSTELALXSLIPTVXXXXXKGTISNLIKIRIKXXVKLPSL 540
DB 481 OSXWKSOMKXISSSTTSXSTELALXSLIPTVXXXXXKGTISNLIKIRIKXXVKLPSL 540
QY 481 OSXWKSOMKXISSSTTSXSTELALXSLIPTVXXXXXKGTISNLIKIRIKXXVKLPSL 540
DB 481 OSXWKSOMKXISSSTTSXSTELALXSLIPTVXXXXXKGTISNLIKIRIKXXVKLPSL 540
QY 541 AFXFONIKXPSIGHLVCTRHCVCHXSMFPMWCSQXFCRRRXLTVVRLFSKNLXTXX 600
DB 541 AFXFONIKXPSIGHLVCTRHCVCHXSMFPMWCSQXFCRRRXLTVVRLFSKNLXTXX 600
QY 541 AFXFONIKXPSIGHLVCTRHCVCHXSMFPMWCSQXFCRRRXLTVVRLFSKNLXTXX 600
DB 541 AFXFONIKXPSIGHLVCTRHCVCHXSMFPMWCSQXFCRRRXLTVVRLFSKNLXTXX 600
QY 601 FNLKRVSRRTXTXITLOISPYHTASTCAXLIPSCVPEFFELSLXXTTPSPHFS 660
DB 601 FNLKRVSRRTXTXITLOISPYHTASTCAXLIPSCVPEFFELSLXXTTPSPHFS 660
QY 601 FNLKRVSRRTXTXITLOISPYHTASTCAXLIPSCVPEFFELSLXXTTPSPHFS 660
DB 601 FNLKRVSRRTXTXITLOISPYHTASTCAXLIPSCVPEFFELSLXXTTPSPHFS 660
QY 661 FFLIVXYITNTCLSEQLIXHRXXSTGEXXLIPIVILALXXAKAGRSJESRVRODPXOGE 720
DB 661 FFLIVXYITNTCLSEQLIXHRXXSTGEXXLIPIVILALXXAKAGRSJESRVRODPXOGE 720
QY 661 FFLIVXYITNTCLSEQLIXHRXXSTGEXXLIPIVILALXXAKAGRSJESRVRODPXOGE 720
DB 661 FFLIVXYITNTCLSEQLIXHRXXSTGEXXLIPIVILALXXAKAGRSJESRVRODPXOGE 720
QY 721 TLSLOKNTKIXPYVLAHTCSLSYSEGKXIXIDXAQVEAAVVRDXALALQPKREERTLS 780
DB 721 TLSLOKNTKIXPYVLAHTCSLSYSEGKXIXIDXAQVEAAVVRDXALALQPKREERTLS 780

```

```

DB 721 TLSLOKNTKIXPYVLAHTCSLSYSEGKXIXIDXAQVEAAVVRDXALALQPKREERTLS 780
QY 781 OKKK 784
DB 781 OKKK 784

RESULT 2
US-10-028-952a-9
Sequence 9, Application US/10028952A
Publication No. US20030157576A1
GENERAL INFORMATION:
APPLICANT: Evin, Jr., Paul R.
TITLE OF INVENTION: EPITHELIAL CELL GROWTH INHIBITORS
FILE REFERENCE: 4273.3USM1
CURRENT APPLICATION NUMBER: US/10/028,952A
PRIOR FILING DATE: 2001-12-18
PRIOR APPLICATION NUMBER: PCT/US00/16900
PRIOR FILING DATE: 2000-06-19
PRIOR APPLICATION NUMBER: 60/139,995
NUMBER OF SEQ ID NOS: 10
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 9
LENGTH: 793
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: UNSURE
LOCATION: (16)..(17)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (42)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (45)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (49)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (65)..(66)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (82)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (99)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (115)..(116)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (122)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (132)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE
LOCATION: (136)
OTHER INFORMATION: Xaa is any amino acid or may be absent
FEATURE:
NAME/KEY: UNSURE

```

LOCATION: (149)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (165) ..(166)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (174)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (179)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (182)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (199)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (215) ..(216)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (232)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (236)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (249)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (265) ..(266)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (274)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (282)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (287)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (299)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (313)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (315) ..(316)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (332)

OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (349)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (365) ..(366)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (380)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (382)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (399)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (415) ..(416)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (432)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (449)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (465) ..(466)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (482)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (499)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (515) ..(516)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (532)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE  
LOCATION: (544)  
OTHER INFORMATION: Xaa is any amino acid or may be absent  
FEATURE:  
NAME/KEY: UNSURE

Query Match 34.3% Score 1288; DB 14; Length 793;  
Best Local Similarity 80.3% Pred. No. 7.6e-122; Indels 4; Gaps 2;  
Matches 282; Conservative 7; Mismatches 58;

Qy 5 HGEETXEVNKKXEPGAPVSPRGARGXRRPCGPVKYHYSDRXXYTDVYRGGEBRGAL 64  
|||||  
Db 37 HGEETXEVNKKXEPGAPVSPRGARGXRRPCGPVKYHYSDRXXYTDVYRGGEBRGAL 96  
|||||  
Qy 65 AXGAKRPAARRPGATNSGXARNGVYLGRTCTQTXQVSGELREDRNLPMKRRARARLI 124

Db 97 ASXARPPAARFGAARSGXAXRWGYLGRYTCQTVXQVSKSELEDRLLPMSXAKARLI 156  
Cy 125 LIFSTNTDXESGASRFXFGFAXXVAKVTGTGTLGRPFXSHDVAFXSDVGSYHXE 184  
Db 157 LIFSTNTDXKSGASRFXFGFAXXVAKVTGTGTLGRPFXSHDVAFXSDVGSYHXE 216  
Cy 185 AEFYKMTIVHPLIGXSWVXTVVRQVSTLLXMCCHGNPAQYERNRRLVYVLGXGA 244  
Db 217 AEFYKMTIVHPLIGXSWVXTVVRQVSTLLXMCCHGNPAQYERNRRLVYVLGXGA 276  
Cy 245 NGAKXSVGLXNLNKSSESPXGTRQRGASVGLGXEPPLSPDAGPPESTXXRAGG 304  
Db 277 NGAKXSVGLXNLNKSSESPXGTRQRGASVGLGXEPPLSPDAGPPESTXXRAGG 336  
Cy 305 RVPRRAPPGSGXAECPSSMETGRG-RKGGXPRLARHAPVRAAEFXXSSTI 354  
Db 337 RAPPRA--GTGVRCXVPFVLNGARPERRPXKRSRPAFGMTWRTXSI 384

## RESULT 3

US-09-925-299-1334  
; Sequence 1334, Application US/09925299  
; Patent No. US20020055627A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA102  
; CURRENT APPLICATION NUMBER: US/09/925, 299  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05883  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1556  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1334  
; LENGTH: 55  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (10)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (49)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (52)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (54)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (55)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-299-1334

Query Match 6.3%; Score 238; DB 9; Length 55;  
Best Local Similarity 86.5%; Pred. No. 2,1e-16;  
Matches 45; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Cy 37 CGPVPYKHYSDRXXTDPVRGGEPRGALAXGAKRPAARRPGATRSXSARWG 88  
Db 2 CGPVPYKHYSDRXXTDPVRGGEPRGALAXGAKRPAARRPGATRSXSARWG 53

## RESULT 4

US-09-925-299-1334  
; Sequence 1334, Application US/09925299  
; Patent No. US20020055627A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA102  
; CURRENT APPLICATION NUMBER: US/09/925,299  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05883  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1556  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1334  
; LENGTH: 55  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (10)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (49)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (52)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (54)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
; NAME/KEY: SITE  
; LOCATION: (55)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-299-1334

Query Match 6.3%; Score 238; DB 10; Length 55;  
Best Local Similarity 86.5%; Pred. No. 2,1e-16;  
Matches 45; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Cy 37 CGPVPYKHYSDRXXTDPVRGGEPRGALAXGAKRPAARRPGATRSXSARWG 88  
Db 2 CGPVPYKHYSDRXXTDPVRGGEPRGALAXGAKRPAARRPGATRSXSARWG 53

## RESULT 5

US-09-925-299-1277  
; Sequence 1277, Application US/09925299  
; Patent No. US20020055627A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
; FILE REFERENCE: PA102  
; CURRENT APPLICATION NUMBER: US/09/925,299  
; PRIOR FILING DATE: 2001-08-10  
; PRIOR APPLICATION NUMBER: PCT/US00/05883  
; PRIOR FILING DATE: 2000-03-08  
; PRIOR APPLICATION NUMBER: 60/124,270  
; PRIOR FILING DATE: 1999-03-12  
; NUMBER OF SEQ ID NOS: 1556  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 1277  
; LENGTH: 40  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: SITE  
; LOCATION: (33)  
; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-299-1277

Query Match 5.3%; Score 199; DB 9; Length 40;  
Best Local Similarity 90.0%; Pred. No. 1,3e-12;  
Matches 36; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Cy 202 WXTTVRQVSTLLXMCCHGNPAQYERNRRLVYVLG 241



Db 1 WYTVVRQVSTFLMMCCCHGNPAQYERNRRFXHLVYVLG 40

RESULT 6  
US-09-925-299-1277

; Sequence 1277, Application US/09925299  
; Publication No. US20030040617A9

; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

; FILE REFERENCE: PA102

; CURRENT APPLICATION NUMBER: US/09/925,299

; PRIOR FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: PCT/US00/05883

; PRIOR FILING DATE: 2000-03-08

; PRIOR APPLICATION NUMBER: 60/124,270

; NUMBER OF SEQ ID NOS: 1556

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1277

; LENGTH: 40

; TYPE: PRT

; ORGANISM: Homo sapiens

; NAME/KEY: SITE

; LOCATION: (33)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-925-299-1277

Query Match

Best Local Similarity 5.3%; Score 199; DB 10; Length 40;

Matches 36; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Db 202 WYTVVRQVSTFLMMCCCHGNPAQYERNRRFXHLVYVLG 241

1 WYTVVRQVSTFLMMCCCHGNPAQYERNRRFXHLVYVLG 40

RESULT 7  
US-09-925-299-1283

; Sequence 1283, Application US/09925299

; Patent No. US20020055627A1

; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

; FILE REFERENCE: PA102

; CURRENT APPLICATION NUMBER: US/09/925,299

; PRIOR FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: PCT/US00/05883

; PRIOR FILING DATE: 2000-03-08

; PRIOR APPLICATION NUMBER: 60/124,270

; NUMBER OF SEQ ID NOS: 1556

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1283

; LENGTH: 91

; TYPE: PRT

; ORGANISM: Homo sapiens

; NAME/KEY: SITE

; LOCATION: (88)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

; NAME/KEY: SITE

; LOCATION: (91)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-925-299-1283

Query Match

Best Local Similarity 5.2%; Score 194.5; DB 9; Length 91;

Matches 42; Conservative 3; Mismatches 15; Indels 11; Gaps 2;

Db 284 PRLSPPAGRP-----PPSTRXKXAGRVPRRPAAGPSGSAECPSWETGGRKGXPL 335

PRLSPPAGRP-----PPSTRXKXAGRVPRRPAAGPSGSAECPSWETGGRKGXPL 335

Db 19 PRLSDSRSPACPRRRAAPPSTRPARAGRVPRRAPGSGAECPSWETGGRKGXPL 75

QY 336 ARAPHVRA 346

Db 76 RLEDPRLTR 86

RESULT 8  
US-09-925-299-1283

; Sequence 1283, Application US/09925299  
; Publication No. US20030040617A9

; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

; FILE REFERENCE: PA102

; CURRENT APPLICATION NUMBER: US/09/925,299

; PRIOR FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: PCT/US00/05883

; PRIOR FILING DATE: 2000-03-08

; PRIOR APPLICATION NUMBER: 60/124,270

; NUMBER OF SEQ ID NOS: 1556

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1283

; LENGTH: 91

; TYPE: PRT

; ORGANISM: Homo sapiens

; NAME/KEY: SITE

; LOCATION: (88)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

; NAME/KEY: SITE

; LOCATION: (91)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-925-299-1283

Query Match

Best Local Similarity 5.2%; Score 194.5; DB 10; Length 91;

Matches 42; Conservative 3; Mismatches 15; Indels 11; Gaps 2;

Db 284 PRLSPPAGRP-----PPSTRXKXAGRVPRRPAAGPSGSAECPSWETGGRKGXPL 335

PRLSPPAGRP-----PPSTRXKXAGRVPRRPAAGPSGSAECPSWETGGRKGXPL 335

Db 19 PRLSDSRSPACPRRRAAPPSTRPARAGRVPRRAPGSGAECPSWETGGRKGXPL 75

QY 336 ARAPHVRA 346

Db 76 RLEDPRLTR 86

RESULT 9  
US-09-925-299-1357

; Sequence 1357, Application US/09925299

; Patent No. US20020055627A1

; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

; FILE REFERENCE: PA102

; CURRENT APPLICATION NUMBER: US/09/925,299

; PRIOR FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: PCT/US00/05883

; PRIOR FILING DATE: 2000-03-08

; PRIOR APPLICATION NUMBER: 60/124,270

; NUMBER OF SEQ ID NOS: 1556

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1357

; LENGTH: 73

; TYPE: PRT

; ORGANISM: Homo sapiens

; NAME/KEY: SITE

; LOCATION: (38)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-925-299-1357

Query Match

Best Local Similarity 5.2%; Score 194.5; DB 9; Length 91;

Matches 42; Conservative 3; Mismatches 15; Indels 11; Gaps 2;

Db 284 PRLSPPAGRP-----PPSTRXKXAGRVPRRPAAGPSGSAECPSWETGGRKGXPL 335

PRLSPPAGRP-----PPSTRXKXAGRVPRRPAAGPSGSAECPSWETGGRKGXPL 335

Db 19 PRLSDSRSPACPRRRAAPPSTRPARAGRVPRRAPGSGAECPSWETGGRKGXPL 75

QY 336 ARAPHVRA 346

Db 76 RLEDPRLTR 86

RESULT 9

US-09-925-299-1357

; Sequence 1357, Application US/09925299

; Patent No. US20020055627A1

; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies

; FILE REFERENCE: PA102

; CURRENT APPLICATION NUMBER: US/09/925,299

; PRIOR FILING DATE: 2001-08-10

; PRIOR APPLICATION NUMBER: PCT/US00/05883

; PRIOR FILING DATE: 2000-03-08

; PRIOR APPLICATION NUMBER: 60/124,270

; NUMBER OF SEQ ID NOS: 1556

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 1357

; LENGTH: 73

; TYPE: PRT

; ORGANISM: Homo sapiens

; NAME/KEY: SITE

; LOCATION: (38)

; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids

US-09-925-299-1357

Query Match

Best Local Similarity 5.2%; Score 194.5; DB 9; Length 91;

Matches 42; Conservative 3; Mismatches 15; Indels 11; Gaps 2;

Db 284 PRLSPPAGRP-----PPSTRXKXAGRVPRRPAAGPSGSAECPSWETGGRKGXPL 335

PRLSPPAGRP-----PPSTRXKXAGRVPRRPAAGPSGSAECPSWETGGRKGXPL 335

Db 19 PRLSDSRSPACPRRRAAPPSTRPARAGRVPRRAPGSGAECPSWETGGRKGXPL 75

QY 336 ARAPHVRA 346

Db 76 RLEDPRLTR 86

NAME/KEY: SITE  
LOCATION: (42)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-299-1357

Query Match 4.9%; Score 185; DB 9; Length 73;  
Best Local Similarity 69.1%; Pred. No. 7.2e-11;  
Matches 38; Conservative 2; Mismatches 7; Indels 8; Gaps 1;

QY 284 PRLSPPRGRP-----PSTRXXRAGGVPRRPAFGSGXACGSPSWETGRGRK 330  
DB 19 PRLASDRSPACPRRRAAFKSTRXARAGGVPRRPAFGSGACGSPSWETGRGRK 73

RESULT 10  
US-09-925-299-1357  
Sequence 1357, Application US/09925299  
Publication No. US20030040617A9  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA102  
CURRENT APPLICATION NUMBER: US/09/925,299  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05883  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1556  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1357  
LENGTH: 73  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (38)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (42)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-299-1357

Query Match 4.9%; Score 185; DB 9; Length 73;  
Best Local Similarity 69.1%; Pred. No. 7.2e-11;  
Matches 38; Conservative 2; Mismatches 7; Indels 8; Gaps 1;

QY 284 PRLSPPRGRP-----PSTRXXRAGGVPRRPAFGSGXACGSPSWETGRGRK 330  
DB 19 PRLASDRSPACPRRRAAFKSTRXARAGGVPRRPAFGSGACGSPSWETGRGRK 73

RESULT 11  
US-10-083-357-1137  
Sequence 1137, Application US/10083357  
Publication No. US20030054370A1  
GENERAL INFORMATION:  
APPLICANT: Qiangdong Zeng et al.  
TITLE OF INVENTION: Systemic Discovery of New Genes  
FILE REFERENCE: 032796-090  
CURRENT APPLICATION NUMBER: US/10/083,357  
CURRENT FILING DATE: 2002-02-27  
NUMBER OF SEQ ID NOS: 1346  
SEQ ID NO 1137  
LENGTH: 57  
TYPE: PRT  
ORGANISM: Saccharomyces cerevisiae  
US-10-083-357-1137

Query Match 4.7%; Score 178; DB 14; Length 57;  
Best Local Similarity 66.1%; Pred. No. 2.8e-10;  
Matches 37; Conservative 4; Mismatches 15; Indels 0; Gaps 0;

QY 117 RRAKARLLIFSTNTDMSGASRSPFPGFYAXXVRRKTTGTTGIMRPSXHSVDVAF 172  
DB 2 KRVKAPLLIFSVNTNHSVAVRSPSPFEPFARCARKVTGTTGIMQSSVHSDIAF 57

RESULT 12  
US-10-083-357-1147  
Sequence 1147, Application US/10083357  
Publication No. US20030054370A1  
GENERAL INFORMATION:  
APPLICANT: Qiangdong Zeng et al.  
TITLE OF INVENTION: Systemic Discovery of New Genes  
FILE REFERENCE: 032796-090  
CURRENT APPLICATION NUMBER: US/10/083,357  
CURRENT FILING DATE: 2002-02-27  
NUMBER OF SEQ ID NOS: 1346  
SEQ ID NO 1147  
LENGTH: 57  
TYPE: PRT  
ORGANISM: Saccharomyces cerevisiae  
US-10-083-357-1147

Query Match 4.6%; Score 174; DB 14; Length 57;  
Best Local Similarity 64.3%; Pred. No. 7.1e-10;  
Matches 36; Conservative 4; Mismatches 16; Indels 0; Gaps 0;

QY 117 RRAKARLLIFSTNTDMSGASRSPFPGFYAXXVRRKTTGTTGIMRPSXHSVDVAF 172  
DB 2 KRVKAPLLIFSVNTNHSVAVRSPSPFEPFARCARKVTGTTGIMQSSVHSDIAF 57

RESULT 13  
US-09-925-299-1314  
Sequence 1314, Application US/09925299  
Patent No. US20020055627A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA102  
CURRENT APPLICATION NUMBER: US/09/925,299  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05883  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1556  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1314  
LENGTH: 54  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-925-299-1314

Query Match 3.9%; Score 145; DB 9; Length 54;  
Best Local Similarity 96.3%; Pred. No. 5.9e-07;  
Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 205 TVRQVSTFLMCCCHGNPAQYERNR 231  
DB 1 TVRQVSTFLMCCCHGNPAQYERNR 27

RESULT 14  
US-09-925-299-1314  
Sequence 1314, Application US/09925299  
Publication No. US20030040617A9  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA102  
CURRENT APPLICATION NUMBER: US/09/925,299  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05883

; PRIOR FILING DATE: 2000-03-08  
 ; PRIOR APPLICATION NUMBER: 60/124,270  
 ; PRIOR FILING DATE: 1999-03-12  
 ; NUMBER OF SEQ ID NOS: 1556  
 ; SOFTWARE: Patent Ver. 2.0  
 ; SEQ ID NO 1314  
 ; LENGTH: 54  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-09-925-289-1314

Query Match 3.9%; Score 145; DB 10; Length 54;  
 Best Local Similarity 96.3%; Pred. No. 5.9e-07;  
 Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 205 TVRQVSFTLLMCCCHGNPAQYERNR 231  
 DB 1 TVRQVSFTLLMCCCHGNPAQYERNR 27

RESULT 15  
 US-09-764-891-4667  
 ; Sequence 4667, Application US/09764891  
 ; Publication No. US2003007808A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rosen et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: PC006  
 ; CURRENT APPLICATION NUMBER: US/09/764,891  
 ; CURRENT FILING DATE: 2001-01-17  
 ; Prior application data removed - consult PALM or file wrapper  
 ; NUMBER OF SEQ ID NOS: 10231  
 ; SOFTWARE: Patent Ver. 2.0  
 ; SEQ ID NO 4667  
 ; LENGTH: 54  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: SITE  
 ; LOCATION: (44)  
 ; OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
 US-09-764-891-4667

Query Match 3.9%; Score 145; DB 10; Length 54;  
 Best Local Similarity 96.3%; Pred. No. 5.9e-07;  
 Matches 26; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 205 TVRQVSFTLLMCCCHGNPAQYERNR 231  
 DB 1 TVRQVSFTLLMCCCHGNPAQYERNR 27

Search completed: April 27, 2004, 11:05:14  
 Job time : 41.2689 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2004, 10:46:49 ; Search time 16.903 Seconds  
(without alignments)  
4461.589 Million cell updates/sec

Title: US-10-028-952A-10  
Perfect score: 3752  
Sequence: 1 IRHGHGHEVKNKKEAPG.....DXAIALQPGXRETLISQKK 784

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	128	3.4	613	4 C40201	artifact-warning s
2	111	3.0	673	4 F40201	artifact-warning s
3	104	2.8	660	1 QOBE3	BHFI1 protein - hu
4	103	2.7	376	2 C75580	adenine deaminase-
5	103	2.7	597	4 E40201	artifact-warning s
6	100	2.7	627	4 A40201	artifact-warning s
7	96	2.6	552	2 F75311	ABC transporter, A
8	95	2.5	705	2 A35363	synapsin I splice
9	94	2.5	1460	1 EDBE1F	immediate-early pr
10	93	2.5	450	2 A38316	alpha-2-adrenergic
11	93	2.5	494	2 G84348	hypothetical prote
12	90.5	2.4	579	4 D40201	artifact-warning s
13	90	2.4	529	2 F45134	hypothetical prote
14	89.5	2.4	769	2 T51112	beta-glucosidase (
15	89.5	2.4	1219	2 S54570	probable membrane
16	88.5	2.4	224	2 E72704	hypothetical prote
17	87	2.3	514	1 W2WLS	E2 protein - human
18	87	2.3	706	2 E30411	synapsin Ia - Dovi
19	86	2.3	373	2 A47234	homeobox protein H
20	86	2.3	450	2 I49481	alpha-2 adrenergic
21	84.5	2.3	2424	2 I46480	calcium channel BI
22	84.5	2.3	464	2 S23697	extensin - Volvox
23	84	2.2	424	2 A54964	spliceosome-associ
24	84	2.2	514	1 W2WLS	E2 protein - human
25	84	2.2	539	2 T28770	hypothetical prote
26	84	2.2	603	2 T45093	hypothetical prote
27	83.5	2.2	282	2 F96687	hypothetical prote
28	83.5	2.2	437	2 S41766	heterogeneous nucl
29	83.5	2.2	684	2 T36771	probable integral

30	83.5	2.2	1215	2 T32734	myosin-IA - Acanth
31	83	2.2	309	2 T47685	probable RNA bindi
32	83	2.2	531	2 T08760	hypothetical prote
33	83	2.2	1106	2 U00405	hypothetical 119.5
34	83	2.2	1248	2 T05059	hypothetical prote
35	82.5	2.2	125	2 A71245	hypothetical prote
36	82.5	2.2	278	2 T46458	hypothetical prote
37	82.5	2.2	380	2 D70516	hypothetical prote
38	82.5	2.2	3345	2 T13423	hypothetical prote
39	82	2.2	312	2 A61183	hypothetical prote
40	82	2.2	344	2 T27410	hypothetical prote
41	82	2.2	704	2 A30411	synapsin Ia - rat
42	82	2.2	1414	1 S23809	collagen alpha 2(I
43	81.5	2.2	414	2 C86301	arginine/serine-ri
44	81.5	2.2	448	2 S05355	hypothetical prote
45	81.5	2.2	839	2 F75518	hypothetical prote

## ALIGNMENTS

RESULT 1  
C40201  
artifact-warning sequence (translated ALU class C) - human  
C/Species: Homo sapiens (man)  
C/Date: 31-Mar-1992 #sequence\_revision 11-Aug-1995 #text\_change 13-May-2000  
C/Accession: C40201  
R/Clavette, J.M.  
Personal communication, 1992  
A/Reference number: A40201  
A/Accession: C40201  
A/Molecule type: DNA  
A/Residues: 1-613 <CIA>  
R/Clavette, J.M.  
Genomics 12, 838-841, 1992  
A/Title: Identifying coding exons by similarity search: Alu-derived and other potential  
A/Reference number: A40200; PMID:92241891; PMID:1572661  
A/Contents: annotation  
C/Comment: This "warning" entry is a conceptual translation in all 6 reading frames of  
in-frame stop codons are shown as 'X'.  
C/Comment: Any significant similarity of a predicted protein sequence to a portion of t

Query Match  
Best local similarity 3.4%; Score 128; DB 4; Length 613;  
Matches 43; Conservative 7; Mismatches 46; Indels 2; Gaps 2;

CY 687 GEXXLPVIALXXAGRSLSRVDPXQHGERTSLQKTKIKXVTLAHCSLSYRSG 746  
DB 2 GVCWMLTPVLPALWEAVTGSXGQEFETSLANNAKPFYKTKKLGHG-AGACSPYALGG 60

CY 747 XGXXIDACEVEAAVRDXAIALQPGXRETLISQKK 784  
DB 61 XGRMAXTRVELAVSR-DLAIALQPGXRETLISQKK 97

RESULT 2  
F40201  
artifact-warning sequence (translated ALU class F) - human  
C/Species: Homo sapiens (man)  
C/Date: 31-Mar-1992 #sequence\_revision 11-Aug-1995 #text\_change 19-May-2000  
C/Accession: F40201  
R/Clavette, J.M.  
Personal communication, 1992  
A/Reference number: A40201  
A/Accession: F40201  
A/Molecule type: DNA  
A/Residues: 1-673 <CIA>  
R/Clavette, J.M.  
Genomics 12, 838-841, 1992  
A/Title: Identifying coding exons by similarity search: Alu-derived and other potential  
A/Reference number: A40200; PMID:92241891; PMID:1572661  
A/Contents: annotation

C/Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c in-frame stop codons are shown as 'X'.

Query Match 3.0%; Score 111; DB 4; Length 673;  
Best Local Similarity 37.2%; Pred. No. 0.011; Indels 2; Gaps 1;  
Matches 35; Conservative 6; Mismatches 51;

691 LIPVILALXXAGXGSLSRVDPQXGHTSLQCKTKIXPYLAFTCSLSYEGXGX 750  
6 LIPVILALMETRRAHLSRGVXDQDQGEFTVSLIKHKLARN--GGPCLSLQHLGROR 63

751 IDXAQVEAAVAXKXALALPQCKRETLISQCKX 784  
64 ENRLNPGXGCEPFLATLQPGCSKGLSKQK 97

## RESULT 3

Q/Species: human herpesvirus 4 (strain B95-8)

C/Date: 25-Feb-1985 #sequence\_revision 25-Feb-1985 #text\_change 23-Aug-1997

C/Accession: A03742

R/Bankier, A.T.; Deininger, P.L.; Farrell, P.J.; Barrell, B.G.

Mol. Biol. Med. 1, 21-45, 1983

A/Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus

A/Reference number: A93065; PMID:85035713; PMID:8092825

A/Accession: A03742

A/Molecule type: DNA

A/Residues: 1-660 <BAN>

R/Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H

Nature 310, 207-211, 1984

A/Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.

A/Reference number: A03794; PMID:84270667; PMID:8087149

A/Contents: annotation; protein coding region

C/Comment: The sequence contains four perfect repeats (residues 149-273, 274-398, 399-52

C/Superfamily: human herpesvirus 4 BHLF1 protein

Query Match 2.8%; Score 104; DB 1; Length 660;  
Best Local Similarity 23.2%; Pred. No. 0.053;  
Matches 87; Conservative 13; Mismatches 103; Indels 172; Gaps 21;

18 APGAP--PVSPRGARCGKRRPCG-----PVYKHYSDRXXTDPVRGGER- 61  
188 APGAPGTPAAP--GPGGAAVPSGATPHERGSGPADPFA-----AARLPERRQ--EPRL 238

62 -GALAYGAKRBAARRGATRGSGXARMGVXLRGTCOTXXQVSGELREDRN---LPVX 116  
239 PDLAALAAQRCPCAG--PPPTSGMAA-----QTHRRPQGPFRSARNPGCPRTWR 285

117 RRAKARLILFTNTDXESGASRFXPGFAXXVAKVTGTLGLMRPSHSDVAFXSD 176  
286 RR-----SGAQGRGHPPPG-----ACQRPSPGPG--GRPAAPG----- 315

177 VGSSTHKEAFTKRWIVHPIIGXSWXIVVRQVSTLLXMCCHGNPNQYERNRXXRL 236  
316 -----APRTPA----- 321

237 VVVLGKGANGAKXXSVGLXNLASKSBRPXGTTORRGASVGLGXFXPLSP----- 288  
322 ----APGPGGAANVPSG-----ATPHPERGSGPADPFAARLPERRQ--EPRL 373

289 PAGRPPSTXXKXAGGVPRRAG-----PGSXAECPSWE-----TGGRK 330  
374 PAG--PPPTSGAAQOTTHRRPPCCPSAKNPG-----CPRTWRSGAGRGHPPRAGQR 427

331 -----GGXPLARNAP 340  
428 PSGETGRPAAPGAP 442

Query Match 2.7%; Score 103; DB 2; Length 376;  
Best Local Similarity 23.0%; Pred. No. 0.037;  
Matches 82; Conservative 15; Mismatches 123; Indels 136; Gaps 18;

20 GAPPVPRGARGGKRRP--CGPVYKHYSDRXXTDPVRGGERGAL---AXGAKRPAAR 74  
64 GRPAPAGAGGQSGAGPFRVAPALY---SGRADPAGAGSAAARVREILGPAPRQD 120

75 RPAATRS--GKSARWGVLAGRYTCOTXXQVSGELR-----EDRNLV---WYRRAKARL 123  
121 RPAVSRRLRRAARLATLRG-----AGVLRRGSGAARRDAPAPARRRRPRARL 169

124 ILFTNTDXESGASRFXPGFAXXVAKVTGTLGLMRPSHSDVAFXSPVGSSTX 183  
170 -----GRHLS-----AALAHADVSPDRH----- 192

184 EAFTKRWIVHPIIGXSWXIVVRQVSTLLXMCCHG---NPAQYERNRXXRLVYVL 240  
193 -----RAGAGGRRRPAAGRRRLRARRVVELL 219

241 GXGANGAKXXSVGLXNLASKSBRPXGTTORRGAS-----VGXGXPL 286  
220 DVGRPARRHP--GHQSHARRASGAPAR--RQRGAPAGSGARTARRRRPGRGRGPRA 276

287 SPFAGRPPTRXKXAGGVPRRAGPGRGSAECPSWETGRGKXGKPLA--RHAP 340  
277 VAPLR--GSDDRPAAGR--GRGCPGPD-----CGGALAGLHLSHHP 318

Query Match 2.7%; Score 103; DB 4; Length 597;  
Best Local Similarity 23.0%; Pred. No. 0.037;  
Matches 82; Conservative 15; Mismatches 123; Indels 136; Gaps 18;

20 GAPPVPRGARGGKRRP--CGPVYKHYSDRXXTDPVRGGERGAL---AXGAKRPAAR 74  
64 GRPAPAGAGGQSGAGPFRVAPALY---SGRADPAGAGSAAARVREILGPAPRQD 120

75 RPAATRS--GKSARWGVLAGRYTCOTXXQVSGELR-----EDRNLV---WYRRAKARL 123  
121 RPAVSRRLRRAARLATLRG-----AGVLRRGSGAARRDAPAPARRRRPRARL 169

124 ILFTNTDXESGASRFXPGFAXXVAKVTGTLGLMRPSHSDVAFXSPVGSSTX 183  
170 -----GRHLS-----AALAHADVSPDRH----- 192

184 EAFTKRWIVHPIIGXSWXIVVRQVSTLLXMCCHG---NPAQYERNRXXRLVYVL 240  
193 -----RAGAGGRRRPAAGRRRLRARRVVELL 219

241 GXGANGAKXXSVGLXNLASKSBRPXGTTORRGAS-----VGXGXPL 286  
220 DVGRPARRHP--GHQSHARRASGAPAR--RQRGAPAGSGARTARRRRPGRGRGPRA 276

287 SPFAGRPPTRXKXAGGVPRRAGPGRGSAECPSWETGRGKXGKPLA--RHAP 340  
277 VAPLR--GSDDRPAAGR--GRGCPGPD-----CGGALAGLHLSHHP 318

Query Match 2.7%; Score 103; DB 4; Length 597;  
Best Local Similarity 23.0%; Pred. No. 0.037;  
Matches 82; Conservative 15; Mismatches 123; Indels 136; Gaps 18;

20 GAPPVPRGARGGKRRP--CGPVYKHYSDRXXTDPVRGGERGAL---AXGAKRPAAR 74  
64 GRPAPAGAGGQSGAGPFRVAPALY---SGRADPAGAGSAAARVREILGPAPRQD 120

75 RPAATRS--GKSARWGVLAGRYTCOTXXQVSGELR-----EDRNLV---WYRRAKARL 123  
121 RPAVSRRLRRAARLATLRG-----AGVLRRGSGAARRDAPAPARRRRPRARL 169

Best Local Similarity 42.9%; Pred. No. 0.06;  
Matches 33; Conservative 10; Mismatches 26; Indels 8; Gaps 4;

QY 711 VRQPPXHGHTLSIQKTKIXPXVLAHTCSLSYSEGGXIXDAQ--EVAABAARXDYAI 768  
DB 226 VXDQDQHGHTPSLTKIXNXL-GVAAHACNPSISGG---XDRRELLSPKOSLRXDCAL 280  
QY 769 ALQPG-XERETLSQKXK 784  
DB 261 VLRRPGQEQNSVSKKK 297

## RESULT 6

A40201  
Artifact-warning sequence (translated Alu class A) - human  
C/Species: Homo sapiens (man)  
C/Date: 31-Mar-1992 #sequence\_revision 11-Aug-1995 #text\_change 19-May-2000  
C/Accession: A40201  
R/Clavette, J.M.  
Personal communication, 1992  
A/Reference number: A40201  
A/Accession: A40201  
A/Molecule type: DNA  
A/Residues: 1-627 <CLA>  
R/Clavette, J.M.  
Genomic 12, 838-841, 1992  
A/Title: Identifying coding exons by similarity search: Alu-derived and other potential  
A/Reference number: A40200; MUID:92241891; PMID:1572661  
A/Contents: annotation  
C/Comment: This "warning" entry is a conceptual translation in all 6 reading frames of c  
in-frame stop codons are shown as 'X'.  
C/Comment: Any significant similarity of a predicted protein sequence to a portion of th

Query Match 2.7%; Score 100; DB 4; Length 627;  
Best Local Similarity 34.2%; Pred. No. 0.12;  
Matches 41; Conservative 10; Mismatches 47; Indels 22; Gaps 3;

QY 673 LSEDLHKKXKXSTGEXKLLPVLIALKXAKGSLSERVADQPHQGETLSLQKTKIXP 732  
DB 92 VSEKXKXKXKXKXAGGLPVLALWEAKAGS-----RQGEITLATTVP 140  
QY 733 XVL-----AHTCSLSYSEGGXIXDAQVEVAABAARXDYALQGERETLSQK 782  
DB 141 RLKXKXKXKXKXAGSAGSPSYLGXGRMAXTRAEALAVR-DRATLQGWROSEPTSRK 199

## RESULT 7

ABC transporter, ATP-binding protein - Deinococcus radiodurans (strain R1)  
F75311  
C/Species: Deinococcus radiodurans  
C/Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
C/Accession: F75311  
R/White, O.; Eisen, J.A.; Heideberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.V.; Lam, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma  
Science 286, 1571-1577, 1999  
A/Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A/Reference number: A75250; MUID:20036896; PMID:10567266  
A/Accession: F75311  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-552 <WHI>  
A/Cross-references: GB:AEO02048; GB:AEO00513; NID:56459929; PIDN:AAFL168.1; PID:5645993  
C/Genetics:  
A/Experimental source: strain R1  
A/Map position: 1

Query Match 2.6%; Score 96; DB 2; Length 552;  
Best Local Similarity 32.1%; Pred. No. 0.27;  
Matches 42; Conservative 7; Mismatches 44; Indels 38; Gaps 9;

QY 228 EBNRXXRLVYVLGXGANGAKXSVGLXLANSKSESRXPXITRQRGASVG--LGXPYR 285  
DB 173 DRAARA-----ACGRGGRTPAAVAGSDARTTPVGR--GTYDRGGLAARLSHPAR 223

QY 286 LSP--AGRPSTEXX-----RAGGRVPR--APGGSXAECPSSWETGGR 329  
DB 224 AAPPDGAC-PRPDAAAPAAARIRHMRPGRPRPRRAAFG-----C-----GNGV 272

QY 330 KGGXPAAHAP 340  
DB 273 GAGDEPARGSP 283

## RESULT 8

A35363  
Synapsin I splice form a - human  
C/Species: Homo sapiens (man)  
C/Date: 17-Aug-1990 #sequence\_revision 17-Aug-1990 #text\_change 01-Dec-2000  
C/Accession: A35363; B35363; A35805  
R/Suedhof, T.C.  
J. Biol. Chem. 265, 7849-7852, 1990  
A/Title: The structure of the human synapsin I gene and protein.  
A/Reference number: A35363; MUID:90243651; PMID:2110562  
A/Accession: A35363  
A/Molecule type: DNA  
A/Residues: 1-705 <SUB>  
A/Cross-references: GB:M58371; GB:J05431  
A/Accession: B35363  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-659; /KASPAQAP<SU2>  
A/Cross-references: GB:M58378; GB:J05431  
R/Sauerwald, A.; Hoese, C.; Oeschwald, R.; Killmann, M.W.  
J. Biol. Chem. 265, 14932-14937, 1990  
A/Title: The 5' flanking region of the synapsin I gene, A G+C-rich, TATA- and CAAT-less  
A/Reference number: A35805; MUID:9038667; PMID:2116513  
A/Accession: A35805  
A/Molecule type: DNA  
A/Residues: 1-125 <SAU>  
A/Cross-references: GB:M55301; NID:938655; PIDN:AAA60608.1; PID:9553654; GB:J05630  
C/Genetics:  
A/Genes: GDB:SYN1  
A/Cross-references: GDB:119606; OMIM:313440  
A/Map position: Xp11.23-Xp11.23  
C/Keywords: actin binding; alternative splicing; phosphoprotein

Query Match 2.5%; Score 95; DB 2; Length 705;

Best Local Similarity 27.5%; Pred. No. 0.43;  
Matches 53; Conservative 16; Mismatches 64; Indels 60; Gaps 11;

QY 177 VGSYHXAETKRWIVHPLIGXS-----WXTVVRQVFTLLXKCCGHNPOYERN 230  
DB 388 VGSXW-----PLIGDHQDEKOLIVLVNKKMOVL-----PRQRD 425

QY 231 RXXRHLYVLGXGANGAKXSVGLXLANSKSESRXPXITRQRGASVGLGXFP-----R 285  
DB 426 AAP-----GKSGQTPSPALPLGQTSQ-QAGSPAAQRRPPQG-GPPQBPBGR 476

QY 286 LSP-AGRP-----STRXXRAGRVPRAPGSAECPSWET-----GRKKG--- 331  
DB 477 GGPPLQGRPPPGQGHLSGLGPPAGSPLFORLSPASPOQASQAPAPPTQCGRQSRV 536

QY 332 ----GXPLARHAP 340  
DB 537 AGPGAPPARAP 549

## RESULT 9

EDB31F  
Immediate-early protein IE180 - suid herpesvirus 1 (strain Indiana-Funkhauser)  
C/Species: suid herpesvirus 1  
C/Date: 30-Jun-1990 #sequence\_revision 30-Jun-1990 #text\_change 16-Feb-1997  
C/Accession: S04713

R:Cheng, A.K.  
Nucleic Acids Res. 17, 4637-4646, 1989  
A:Title: DNA nucleotide sequence analysis of the immediate-early gene of pseudorabies virus  
A:Reference number: 504713; MUID:89315207; PMID:2346124  
A:Accession: S04713  
A:Molecule type: DNA  
A:Residues: 1-1460 <CHE>  
C:Superfamily: herpesvirus immediate-early protein 1E175  
C:Keywords: DNA binding; early protein; transcription regulation

Query Match 2.5%; Score 94; DB 1; Length 1460;  
Best Local Similarity 20.8%; Pred. No. 1.2;  
Matches 73; Conservative 18; Mismatches 122; Indels 138; Gaps 15;

```

QY 18 APGAPVSPRGARG-----GXRPPCPVRYHYSDRXXTDPVRGQ-----58
DB 184 APAPAP-APRRPRDGGPRGRTSVSPGRRRGLGPRRHGHSQGRW---PGRHGGGFLPQ 239
QY 59 --EPRGALAGAKRPAARPGATRSGXSAWGVXGRTQCTXXQVSGELREDRLPMX 116
DB 240 PPPPPGR-----SRPPAAAPPPAE-----GTAIVTITSTAS-----PWL 274
QY 117 RRAARLLIFSTMTDSSGASRSFPGFXAXXVRAVTTGITGLMRPSXHSVAFXSFD 176
DB 275 -----DBPAAARLDP-----AAARPEER-----294
QY 177 VGSYXHEAEFTKMTIYHPLIGXXSWYTVVRQVSTLLXMCCHGNPAQYERNRXXRL 236
DB 295 -----LQDP-----QLQLD-----HHRRRRRAR 315
QY 237 VYVUGXGNGAKXXSVGLXMASSESRRPGTIRORRG-ASVGLGXPRLSPPAGRP 295
DB 316 RPRRRGRTPRRGSGADLQROPRRRAGEGALRGKRFSSSSGSDSLSP--ASPS 373
QY 296 STRXXRAGRVRRAPGPGSXAECPS-----WETGRKRGKXPLARHAP 340
DB 374 APRAPAAAAAARSASSSSSSSSSSSSSEGEDEGVAPGAPLARAGP 424

```

## RESULT 10

A38316  
alpha-2-adrenergic receptor - pig  
C:Species: Sus scrofa domestica (domestic pig)  
C>Date: 22-Jan-1993 #sequence\_revision 22-Jan-1993 #text\_change 13-Aug-1999  
R:Guyer, C.A.; Hotzman, D.A.; Wilson, A.L.; Clark, J.D.; Cragoe Jr., E.J.; Limbird, L.E.  
J. Biol. Chem. 265, 17307-17317, 1990  
A:Title: Cloning, sequencing, and expression of the gene encoding the porcine alpha-2-adrenergic receptor  
A:Reference number: A38316; MUID:91009167; PMID:2170371  
A:Accession: A38316  
A:Molecule type: DNA  
A:Residues: 1-450 <GV>  
A:Cross-references: GB:J05652; NID:g164303; PIDN:AAA30984.1; PID:g164304  
C:Superfamily: vertebrate rhodopsin  
C:Keywords: G protein-coupled receptor; glycoprotein; transmembrane protein

Query Match 2.5%; Score 93; DB 2; Length 450;  
Best Local Similarity 24.3%; Pred. No. 0.43;  
Matches 43; Conservative 14; Mismatches 74; Indels 46; Gaps 8;

```

QY 241 GKGANGAKXXSVGLXMASSESRRPGTIRORRGASVGLGXPRLSPPARPPSRXX 300
DB 258 GVGRRGAEAPLPVQJNGAPGPAPAGP---RQAD-GLDUESSSSSHARPPGPRHSE 312
QY 301 RA-----GGRVRR---APGGSXACPSWETGRKRGKXPLARHAPVR 343
DB 313 RGPRAKSKARASQVVRGDSLPFRGPGAGPGAPA-----TGAGEERG-----VA 357
QY 344 AADEFXXSTTNRHTSACIFKXXILFLW-----VDIQKWDCK---XTXKFTYWF 391
DB 358 KASRNRRGRRREKFTFLAVVIGVFVVCWPFPFYTTTLTAVGCVSPPTLRFKFFWF 414

```

## RESULT 11

G84348  
hypothetical protein Vng1983h [imported] - Halobacterium sp. NRC-1  
C:Species: Halobacterium sp. NRC-1  
C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Feb-2001  
A:Accession: G84348  
A:Residues: 1-494 <STO>  
A:Reference number: A84160; MUID:20504483; PMID:11016950  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-494 <STO>  
A:Cross-references: GB:A8004437; NID:g10581414; PIDN:AMG20155.1; GSPDB:GN00138  
C:gene: VNG1983h

Query Match 2.5%; Score 93; DB 2; Length 494;  
Best Local Similarity 24.8%; Pred. No. 0.47;  
Matches 102; Conservative 26; Mismatches 183; Indels 100; Gaps 24;

```

QY 1 ITHHEGETXEVXNKEAPGAPVSPRGARGXR-PCGPPYKYHYSDRXXTDPVRGGE 59
DB 62 LTPARGTGVDAAEDDAPVPTVADGGGQTRCPSPGVSQVT---AEPVRR--R 116
QY 60 PRGALAXGAKRPAAR--REGATR-----SGXSARNG-----VXLGRTCTQXXQ 101
DB 117 PVGRLLRVLRGRAVDRAVGRARADTASGRRPRGRTPHRADVCGCGPVD--R 173
QY 102 VEXGELREDRN-LPMXRRAKRLILIFST---NTDXSGASRSFPGFXAXXVRKYTT 156
DB 174 VGANRPAEPRRLHCHCRRLRRLVALQDALGRAGDRHGA-----VLPBVRLLVP 225
QY 157 GITGL-----WPSXHSDVAFYSPDVG-SSYXHEAEFTKMTIYHPLIG-----XX 200
DB 226 RVGRPVAVDLRPPGRDTR-PALDPGRTAADRRPRPR---RPPVGGPRRPLVRPLR 281
QY 201 SWXTVVRQVSTLLXMCCHG-----NPAQYERNRXXHLYVVGXGANGAKXXSV 252
DB 282 HLYDRIPRCRGRSLVAAARHVVGGDSAPRGAAARRPRRGR-----GLAGARQGV 334
QY 253 -GLXMASSE--SRPGTIRORRGASVGLGXPR-RUSPPAGRPSPS---TRXXRAG 304
DB 335 PGRRRARGLRLAARAAVAGRGHGFGLGRRPARLDDPGRRPAPVRYDSRRRTRG 394
QY 305 ---RVRRAPGSGSXAECSSWETGRKRGKXPLA-----RHAPHYRAR 345
DB 395 RRRRPPORARRPGARQC-----RPSKALPLAVTGRPRGRSLAGTRAR 438

```

## RESULT 12

A40201  
artifical-warming sequence (translated Alu class D) - human  
C:Species: Homo sapiens (man)  
C>Date: 31-Mar-1992 #sequence\_revision 11-Aug-1995 #text\_change 19-May-2000  
A:Accession: A40201  
A:Reference number: A40201  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-579 <CIA>  
R:Claverie, J.M.  
Genomics 12, 838-841, 1992  
A:Title: Identifying coding exons by similarity search: Alu-derived and other potential  
A:Reference number: A40200; MUID:92241891; PMID:1572661  
A:Contents: annotation  
C:Comment: this "warning" entry is a conceptual translation in all 6 reading frames of in-frame stop codons are shown as 'X'.



C:Comment: Any significant similarity of a predicted protein sequence to a portion of th

```

Query Match      2.4%; Score 90.5; DB 4; Length 579;
Best Local Similarity 35.2%; Pred. No. 0.98;
Matches 31; Conservative 9; Mismatches 33; Indels 15; Gaps 4;

QY 701 AKAGRSLESPRDPXOHGETLSIQ-KNTKIX-----PYVLAHTCSLSYSEGGXIXDX 753
    15 AMAGVSSSGVQEPGQHVETPLQKDTISQWKHKVVPVPA-----TKAGAGSLDI 69
    754 AQEVEAAAVRXDAIALQPKXRETTISQ 781
    70 GSQ---GSVSHDTSVLPQGHKTLP 94

Db

RESULT 13
T51112
hypothetical protein [imported] - Microbacterium ammoniaphilum (fragment)
C:Species: Microbacterium ammoniaphilum
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T51112
R:Strickland, H.M.; Seeber, S.; Jarsch, M.; Kessler, C.
Gene 172, 41-46, 1996
A:Title: Cloning and characterization of the ManI restriction-modification system from M
A:Reference number: 222923; MUID:96257250; PMID:8654988
A:Accession: T51114
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-529 <STR>
A:Cross-references: EMBL:X79027; NID:984667; PIDN:CA55649.1; PID:g1679831
A:Experimental source: ATCC 15354

Query Match      2.4%; Score 90; DB 2; Length 529;
Best Local Similarity 21.9%; Pred. No. 1;
Matches 87; Conservative 23; Mismatches 140; Indels 148; Gaps 18;

QY 2 RHHEGEETXEV---XNKKEAAGAP---PVSFRGARGXRRPCGP---PVKTHYSDRXXTD 52
    70 RHRRPAQRQVHPLQRPHEGARGELFVRDRAERGRREPAPARDARHL--RUGAH 127
    53 FVRRG-----GE-----FRGALXGAXRPARRRPGATRSXSARWG 88
    128 PARQGVLRHRRHARGRGRGPRQVPRQHPRG-----RRDAPGRGLHRRRRAR-- 179
    QY 89 VXLGRYTCQYXQVSG--ELREDRNLFWKRAARLILITSTYDSEGSRSFXFGF 146
    180 -----RRQGRFQVHGDQHRADFRPRRPRAGHNP-----RGGQGGEARPL-- 223
    QY 147 XAXXVRKVTGTLGLMRPSXHSDDVAFXSFDVGSSTYHAEFTKEMIVHPLIGXSWXTV 206
    224 -----PARGGQGGGPRGR-----HPALGD----- 244
    QY 207 VRQVSTLLXMCCHGNAPQYERNRKHLYVTGKGANGKXXSVGLXLANASSES--- 263
    245 -----EDRRARQ-----GAFAHRAALHLRLQRRGRADGC 276
    QY 264 RPXGTRRQRGASVGLXPRSLPPAGRPP-----STRXXRG-----RVP 307
    277 RAQGAARRARRRGGG-GAPREDRVRAIRPRRGRGGAAALDRAGVGARPARRRRLRHP 335
    QY 308 RRAFGGSAECPSMETGRGRKGGPLARHAPVRRAR 345
    336 RPADLPFGAGPGQVARDLHPQGLEGAAGSRHRHRLRAR 373

Db

RESULT 14
T51112
beta-glucosidase (EC 3.2.1.21) oler [validated] - Streptomyces antibioticus (ATCC 11891)
C:Species: Streptomyces antibioticus
A:Variety: ATCC 11891
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C:Accession: T51112

```

R:Quiroz, L.M.; Aguirrezabalaga, I.; Olano, C.; Mendez, C.; Salas, J.A.  
submitted to the EMBL Data Library, September 1999

```

A:Reference number: 225300
A:Accession: T51112
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-769 <QUT>
A:Cross-references: EMBL:AF055579; PIDN:AA012650.1
A:Experimental source: ATCC 11891
C:Genetics:
A:Gene: oler
C:Function:
A:Description: extracellular reactivation of oleandomycin; converts glycosylated oleam
C:Keywords: glucosidase; hydrolase

Query Match      2.4%; Score 89.5; DB 2; Length 769;
Best Local Similarity 29.1%; Pred. No. 1.7;
Matches 34; Conservative 8; Mismatches 48; Indels 27; Gaps 4;

QY 237 VYTLGXGANGAKKXSVGLXANASKESRXPXGTRRGA-----SVC---LGXPRLS 287
    334 IAVIGPTAKVPKVTGIGSSYIVPDGASAPLDTRRAGAGSTRVSTGBETVGVFPQ-S 392
    QY 288 PRAGRPPESTRXXRAGC-----RVPRAPPGSGXACPSMETGR 327
    393 ALPRPRPBGVFPAGCGVLYDGTITVYVGSYRIARAGGNAYVELDQGEPRGR 449
    Db

RESULT 15
S54570
probable membrane protein YMR109w - Yeast (Saccharomyces cerevisiae)
N:Alternate names: hypothetical protein YM9718.08
C:Species: Saccharomyces cerevisiae
C:Date: 08-Jul-1995 #sequence_revision 19-Oct-1995 #text_change 02-Feb-2001
C:Accession: S54570
R:Hunt, S.; Bowman, S.
Submitted to the EMBL Data Library, May 1995
A:Reference number: S54510
A:Accession: S54570
A:Molecule type: DNA
A:Residues: 1-1219 <HUN>
A:Cross-references: EMBL:249702; NID:g817859; PID:g817867; MIPS:YMR109w
A:Experimental source: strain Ab972
C:Genetics:
A:Gene: SGD:MYO5
A:Map position: 13R
A:Cross-references: SGD:S0004715; MIPS:YMR109w
C:Superfamily: protozoan myosin heavy chain IB; myosin motor domain homology; SH3 homo;
C:Keywords: nucleotide binding; P-loop; transmembrane protein
F:39-702/Domain: myosin motor domain homology <MMOT>
F:129-136/Region: nucleotide-binding motif A (P-loop)
F:293-309/Domain: transmembrane #status predicted <TMW>
F:1092-1142/Domain: SH3 homology <SH3>

Query Match      2.4%; Score 89.5; DB 2; Length 1219;
Best Local Similarity 26.8%; Pred. No. 2.7;
Matches 30; Conservative 11; Mismatches 44; Indels 27; Gaps 4;

QY 256 INASKESRPPXGTRR-----GASVGLXKPYRFLSPAGRPSTRTXXRAGRV 307
    1035 LTNSGNSAPRPTATPATPATPATPAAMGSGRQA-NIPPPPPPPSPSKPEPEFEAA 1093
    QY 308 RRAFGGSAEC-----PSMETGR-----GRKGGXPLARHAP 341
    1094 YDFPGSGSPBELPLKKGDIYITREPSGMSKGLDGSKEGVNPTANMKPH 1145
    Db

```

Search completed: April 27, 2004, 10:56:03  
Job time : 18.903 secs